In accordance with the latest research, this book presents several segments that describe the direction and techniques together with actual implementation of new methods for controlling and improving of the quality of services in future generation mobile networks. It gives review of the architectures of the future generations of networks, it performs analysis of network architectures in heterogeneous networks and the ways radio resource management is achieved in each of them. The main focus was given on the heterogeneous wireless networks, through which a new proposal is presented for their architecture of internetwork operation using the Internet model for interoperability between radio access technologies that participate in the construction of heterogeneous networks. The proposed new architecture in this book defines the unique ecosystem of inter-network operation between different RAT’s that form next generation mobile networks. Furthermore, the book defines a new algorithm for selecting of access network in heterogeneous wireless environment by using algorithms from artificial intelligence. This proposal can be base for implementation of new 5G architecture.

Aleksandar Tudzarov

Author received his M.Sc. and Ph.D. degrees in Electrical Engineering from Ss. Cyril and Methodius University in Skopje. Since 2012, he has been an assistant professor on the Faculty of Electrical Engineering at Goce Delcev University in Shtip, R.Macedonia. His main work is in the area of mobile and wireless packet-based networks and QoS.

978-3-8433-7765-2

QoS in NGN mobile and wireless networks

Tudzarov

Aleksandar Tudzarov

Quality of Service in next generation mobile and wireless networks

Way towards new service based 5G architecture

LAP LAMBERT Academic Publishing

Aleksandar Tudzarov

Quality of Service in next generation mobile and wireless networks

Aleksandar Tudzarov

Quality of Service in next generation mobile and wireless networks

Way towards new service based 5G architecture

LAP LAMBERT Academic Publishing

Impressum / Imprint
Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
Alle in diesem Buch genannten Marken und Produktnamen unterliegen warenzeichen-, marken- oder patentrechtlichem Schutz bzw. sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Inhaber. Die Wiedergabe von Marken, Produktnamen, Gebrauchsnamen, Handelsnamen, Warenbezeichnungen u.s.w. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutzgesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.
Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this works is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Coverbild / Cover image: www.ingimage.com

Verlag / Publisher:
LAP LAMBERT Academic Publishing
ist ein Imprint der / is a trademark of
AV Akademikerverlag GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Germany
Email: info@lap-publishing.com

Herstellung: siehe letzte Seite /
Printed at: see last page
ISBN: 978-3-8433-7765-2

Zugl. / Approved by: Skopje, University of Ss. Cyril and Methodius , Diss., 2011

CONTENTS

LIST OF ACRONYMS

2G	Second Generation of Mobile Networks
3G	Third Generation of Mobile Networks
3GPP	3rd Generation Partnership Project
4G, 5G	Fourth, Fifth Generation of Mobile Networks
AAA	Authentication Authorization and Accounting
ABC	Always Best Connected
ACK	Acknowledge
AH	Authentication Header
AP	Access Point
API	Application Programming Interface
AVP	Attribute Value Pairs
BDP	Bandwidth delay product
BoA	Bisector of Area
BOOTP	Bootstrap Protocol
CCA	Charging Control Application
CDF	Cumulative Distribution Function
CDMA	Code division multiples access
CFP	Contention Free Period
CHAP	Challenge Handshake Authentication Protocol
CoA	Centroid of Area
CPE	Customer Premises Equipment
CPH	Client Profile Handler
CQPBR	Central Quality Policy Based Router
CTS	Clear To Send
DCF	Distributed Coordination Function
DHCP	Dynamic Host Configuration Protocol
DIFS	Distributed Interframe Space
DNS	Domain Name System
DVB	Digital Video Broadcast
EAP	Extensible Authentication Protocol
EAP- MD5	EAP Message Digest 5
EAP-OTP	EAP One Time Password
EAP-SIM	EAP Subscriber Identification Module

EAP-TLS	EAP Transport Layer Security
EAP-TTLS	EAP Tunneled TLS Authentication Protocol
EAP-SAML	EAP Security Assertion Markup Language
EC	European Commission
EDGE	Enhanced Data Rates for GSM Evolution
EP	Enforcement Point
ePDG	Evolved Packet Data Gateway
EPS	Evolved Packet System
ESP	Encapsulation Security Payload
FCS	Frame Check Sequence
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FPGA	Field-programmable gate array
FWA	Fixed Wireless Access
GA	Genetic Algorithm
GPRS	General Packet Radio System
GRE	Generic Routing Encapsulation
GSM	Global System of Mobile
GTP	GPRS Tunneling protocol
IAS	Internet Authentication Services
ICMP	Internet Control Message Protocol
ICMP	Internet Control Message Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IKE	Internet Key Exchange
IPSec	IP Security
ITHC	Inter Tunnel Handover Control
KPI	Key Performance Indicators
L2TP	Layer 2 Tunneling Protocol
LEAP	Lightweight EAP
LNS	L2TP Network Server
LOM	Largest of Maximum
LRD	Long Range Dependency
LTE	Long Term Evolution
MAC	Medium Access Control
MAP	Mobile Application Part

MCCSM	Media Connection Control Software Module
MCDM	Multi-Criteria Decision Making
McIP	Mobile client IP Address
MIM	Mobile Identity Management
ML	Maximum Likelihood
MPDU	MAC Protocol Data Unit
MQPBR	Mobile Quality Policy Based Router
MS-CHAP	Microsoft – CHAP
MTU	Maximum Transmission Unit
NAS	Network Access Server
NAT	Network Address Translation
NAV	Network Allocation Vector
NFS	Network File System
NIC	Network Information Center
OCI	Open Charging Interface
OFDM	Orthogonal Frequency Division Multiplex
OSI	Open Systems Interconnection model
OTP	One Time Password
PAA	PANA Authentication Agent
PaC	PANA Authentication Client
PANA	Protocol for carrying Authentication for Network Access
PAP	Cleartext Password Authentication Protocol)
PCF	Point Coordination Function
PDA	Personal Digital Assistant
PDF	Probability Density Function
PDP	Packet Data Protocol
PEAP	Protected EAP
PHY	Physical Layer
PKI	Public Key Infrastructure
PLCP	Physical Layer Convergence Protocol
PLMN	Public Land Mobile Network
PMD	Physical Medium Dependent Sublayer
PME-FE	Performance measurement execution functional entity
PMP-FE	Performance measurement processing functional entity
PMR-FE	Performance measurement reporting functional entity
PPP	Point to Point Protocol

PPTP	Point to Point Tunneling Protocol
PR	Price
PSO	Particle Swarm Optimization
PBR	Policy Based Router
QoE	Quality of Experience
QoS	Quality of Service
QoS/E CM	QoS and QoE Control Manager
QoSPRO	Quality of Service Policy based Routing
RACF	Resource and Admission Control Function
RADIUS	Remote Dial-In Remote Access Service
RAT	Radio Access Technology
RAT-CCSM	RAT – Connection Control Software Module
RFC	Request for Comment
RLC	Radio Link Control
RRM	Radio Resource Management
RTCP	Real Time transport Control Protocol
RTCP	Real Time Control Protocol
RTP	Real Time transport Protocol
RTS	Request To Send
RTT	Round Trip Time
RUNE	Rudimentary Network Emulator
RTP	Real Time Protocol
SA	Security Associations
SADB	Security Association Database
SCTP	Stream Control Transport Protocol
SDR	Software Defined Radio
SIM	Subscriber Identity Module
SL	Signal Level
SMART	Simple Multi-Attribute Rating Technique
SOM	Smallest of Maximum
SPI	Security Parameter Index
SPME	Security and Policy Management Entity
STA	Station
TACACS	Terminal Access Controller Access Control System
TDD	Time Division Duplex
TDMA	Time Division Multiple Access

TEID	Tunnel Endpoint Identifier
TIM	Traffic Indication Map
TKIP	Temporal Key Integrity Protocol
TLS	Transport Layer Security
ToS	Type of Service
TS	Type of Service
TTL	Time To Live
TV	Terminal Velocity
UAM	Universal Access Method
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
VLAN	Virtual LAN
VPN	Virtual Private Network
WCDMA	Wideband Code division multiples access
WEP	Wired Equivalent Policy
WISP	Wireless Internet Service Provider
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WPA	Wi-Fi Protected Access
WPAN	Wireless Personal Area Network
WWAN	Wireless Wide Area Network
WWW	World Wide Web
XDR	External Data Representation
SNMP	Simple Network Manage Protocol
SMTP	Simple Mail Transport Protocol
MSE	Minimum Square Error
HTTP	HyperText Transfer Protocol
TFTP	Trivial File Transfer Protocol
TCP	Transmission Control Protocol

PREFACE

Today we have different wireless and mobile technologies, which are mass deployed, such as 3G mobile networks (UMTS, cdma2000), LTE (Long Term Evolution), WiFi (IEEE 802.11 wireless networks), WiMAX (IEEE 802.16 wireless and mobile networks), as well as accompanying networks, such as sensor networks, or personal area networks (e.g., Bluetooth, ZigBee). Mobile terminals include variety of interfaces, including the GSM ones, which are based on old-fashioned circuit switching, the technology that is going into its last decade of existence. All wireless and mobile networks today are going towards all-IP principle, meaning all data and signaling will be transferred via IP (Internet Protocol) on network layer. With the continued development and growth of wireless networks in terms of implementation of new technological systems, number of subscribers and in terms of new and innovative services based on existing technology, the need to find a successful way of controlling and improving quality of services offered to end users appears. The consumer community now demands evermore powerful functionality and continuously improving applications not only by offering new and innovative functionalities but also by improving its quality.

This new and more sophisticated user demand is driving the research community to look toward the future of quality of service definitions in next generation wireless networks. The proper way of determining of offered quality of services is crucial in analogy with real radio conditions in a wireless network. Providing quality customer service which ensures the fulfillment of user requirements is imperative for any wireless carrier. To achieve the objectives in order to provide required service quality, proper definition and continuous monitoring of the services is mandatory. Quality of service offered to the user are defined on the basis of user requirements and network performance, and accordingly, compliance with these requirements need to be monitored exactly where quality of service is required, at the user side. This definition actually requires service monitoring by the user, not by the network that provides the service. The introduction of continuous monitoring of customer service as the solution enables emulation of customer services as well as the establishment of a system for connecting parts of the service to specific points in the network infrastructure of the operator. In this way full performance monitoring of individual network elements and segments that are involved in the realization and maintenance of the requested service is realized.

The aim of this book is to meet described challenges. In the book, we give an overview of the characteristics of data services and their specifics related to wireless and mobile systems. It presents new network architecture for next generation wireless networks beyond 4G, with special attention on providing QoS to the end user.

Next generation of mobile and wireless networks will certainly need to fit within the NGN, because it is based on wireless and wired access possibilities, including all services and using all-IP concept. However, the main principle for NGN is complete separation in parallel between the transport part in the access and in the core networks from the service provisioning, i.e., from the service stratum. Since, the 4G is already at the “front door” of communication world, the next generation of mobile and wireless networks will be labeled 5G, if we continue the same pattern from the past two decades. We believe that the 5G approach will be user-centric approach, since the mobile terminals are becoming highly computationally capable devices which can support more complex functionalities for performing calculations, as well as bigger memory space and longer battery life.

It discuss possible ways to determine and define the parameters that give a clear picture of the quality of services given to the end user, as well as their connection with QoS parameters defined at the network level. It gives overview of the characteristics of mobile and wireless Internet services and their critical parameters in terms of quality as well as the proposal for new methods to control the QoS according to the design of future generations of mobile networks, which will successfully analyze network performance, reflected by offered quality customer service. Furthermore, the book defines a new algorithm for selecting of access network in heterogeneous wireless environment by using algorithms from artificial intelligence. Presented algorithm, called "M-RATS", was placed in Simulation environment and an analysis of its performance in terms of customer satisfaction was made.

The functionality of the algorithm and its construction has made quite flexible and usable procedures for network selection and the initial choice of access technology. The system is applicable in environments where there is a need to adopt a decision based on diverse parameters and criteria for making decisions based on their previous history.

The proposed new solutions and innovations presented in the book give an important contribution to future generations of mobile networks, where customer service will be separated from transport technologies and customer satisfaction of various services in heterogeneous wireless and mobile environment will be their primary goal.

Chapter I

INTRODUCTION

The rapid development of telecommunications which has intensified in the last two decades, as well as need for generalization of telecommunications networks in order to offer service integration, results in continuous improvement of network architectures and mechanisms in order to meet user requirements. In this context and due to particularly intense and continuous pressure by the users and their service requirements, development of next generation wireless networks whose architecture should support existing and provide quick and easy development of new services has been initiated. When it comes to next-generation wireless networks, in its essence the system consists of wireless systems providing a continuum of services to the user. The objective of this book is to cover a significant segment of future network architectures, particularly related to providing quality of services and mechanisms for control and continuous monitoring in order to ensure satisfactory level of perceptive quality of services offered to end-user. This book makes an overview of several procedures and methods for analysis of customer service quality of IP-based services in wireless and mobile networks, with particular reference to their performance and their relationship with the network parameters, and the impact of network parameters on performance and presentation quality.

Within the book, new methods were developed that should contribute to effective analysis of the performance of mobile network offered in terms of quality of customer services. Based on the analysis and created systems for quality control of services a detail analysis to determine the relationship between IP packet services and QoS parameters defined at the network level was conducted as well as detailed analyses of relationship with the bearer services on mobile and wireless networks. In addition, new solutions for improvement of the quality of customer service offered in heterogeneous mobile and wireless networks were designed. In the first chapter of the book, an overview of the concept of heterogeneous networks and their representation of the mobile terminal and the transmission network is presented. The description of the models for interoperability among heterogeneous networks is presented initially and then a proposal for network architecture for interoperability between systems in the future generation of wireless networks was introduced. In the following part of the book the basic use cases in the proposed architecture are presented and design of the proposed network-level connection, transparent to higher network levels, through heterogeneous networks is given. The use cases analyzed the processes and mechanisms for authentication, the concept of providing application QoS, thus defining and applying the policy based routing in the proposed architecture. In this section basic qualitative parameters of packet based services (IP-based) are analyzed and give an overview of the defined KPI (Key performance Indicators). To enable more comprehensive analysis, a measurement system for quality control of user services that identifies individual aspects of the service is developed and new techniques that measure the individual performance of the system are introduced.

The text covers radio resource management in heterogeneous wireless networks in terms of access network selection. It defines mechanisms for single and joint radio resource management and at the same time defines the levels of interconnection between different radio access technologies within the heterogeneous wireless networks. This section defines the types and mechanisms for selecting the access network.

The third chapter includes a new set of algorithms for selecting an access network based mechanisms inspired by nature. Construction of new algorithm for access network selection is presented. The inspiration for designing a new algorithm is found in the mechanisms and software modules based on artificial intelligence and algorithms inspired by nature. Based on these mechanisms, intelligent algorithm is designed which refers to a heterogeneous wireless network and covers all radio access technologies (networks). The solution for access network selection which is based on this design, selects the best available technologies and ranks access technologies according to their acceptability and dutifulness regarding operator and user preferences simultaneously.

In the forth chapter an evaluation of the proposed algorithm for selecting the access network through simulation and analysis of the results is given. This is a review of the simulation model and methods of joint work of the optimization algorithms and their configuration within the proposed simulation. In this chapter, evaluation of the performance of the proposed algorithm is presented. The evaluation clearly shows that the proposed (M-RATS) algorithm gives the best results for quality of service from user perspective compared to other algorithms for selecting of the access network through simulation and analysis of the results.

Chapter II

CONCEPT OF HETEROGENOUS NETOWRKS AND THEIR INTRODUCTION ON MOBILE TERMINAL SIDE OVER RADIO ACCESS NETWORKS

Introduction

Today in the world we have widely deployed different wireless and mobile technologies, such as 3G mobile networks (UMTS, cdma2000), LTE (Long Term Evolution), WiFi (IEEE 802.11 wireless networks), WiMAX (IEEE 802.16 wireless and mobile networks), as well as accompanying networks, such as sensor networks, or personal area networks (e.g., Bluetooth, ZigBee). Mobile terminals include variety of interfaces, including the GSM ones, which are based on old-fashioned circuit switching, the technology that is going into its last decade of existence. All wireless and mobile networks today are going towards all-IP principle, meaning all data and signaling will be transferred via IP (Internet Protocol) on network layer [1].

Thus, we may have different Radio Access Technologies (RATs) today and new RATs in the future (e.g., LTE-Advanced), but the common “thing” for all of them is IP, which is unifying technology. The 4G term is related to available bit-rates in the access link, i.e. more than 1 Gbps is set as condition by ITU for a technology to be marked as 4G. Also, all-IP is the characteristic of 4G in the access and in the core network part, and there will be no circuit-switching as in 3G systems, such as UMTS. On the other side, lot of efforts are being done for separation of transport stratum and service stratum in the concepts of Next Generation Networks (NGN), [2], [3]. Next generation of mobile and wireless networks will certainly need to fit within the NGN, because it is based on wireless and wired access possibilities, including all services and using all-IP concept. However, the main principle for NGN is complete separation between the transport part in the access and in the core networks from the service provisioning in parallel, i.e. from the service stratum. Since, the 4G is already at the “front door” of communication world, the next generation of mobile and wireless networks will be labeled 5G, if the same pattern from the past two decades continues. We believe that the 5G approach will be user-centric approach [4], since the mobile terminals are becoming highly computationally capable devices which can support more complex functionalities for performing calculations, as well as bigger memory space and longer battery life in years will provide enough storage capability for control information. In the IP world, the main principle since the beginning was keeping simple network nodes and having smart end devices (e.g., computers), an approach completely different from the Public Old Telephone Systems (POTS). However, we need smart nodes on the networks side in all-IP concept as well, which should be used for negotiation with the user equipment premises (mobile terminals in the case of mobile networks) for providing essential Quality of Service, as well as authentication, authorization, accounting and security functionalities.

This chapter provides complete functional architecture for next 5G mobile networks. The main assumption in our approach is that the user will have the possibility to access different RATs from single mobile device at the same time, which is reality even today. Furthermore, we propose establishing new network nodes for policy-based routing between IP tunnels to mobile user via different RATs, which are placed in service stratum of the network. We have invented several solutions for making the proposed 5G network architecture fully functional.

Interoperability in Heterogeneous Wireless Environment

The challenge in the design of the terminals is connected with the management of trade between the flexibility of how to use the spectrum and needed space and power to given platform. New methods for partial reconfiguration offer design dimensions that allow the system to adapt to the opportunities and requirements of the terminals in a manner that shall maximize the spectral efficiency and maximize the battery power as well [5]. As a result of growing level of acceptance of the wireless technologies in different fields, the challenges and types of wireless systems associated with them are changing.

With the evolution of 3G/4G cellular systems defined by 3GPP, new architecture provides sophisticated control mechanisms that enable the central management of the operator's network in granular way with great precision and accuracy. In this context 3GPP introduced new methods for providing radio management (hierarchical management of resources) implemented in systems with a common radio resource management, a single radio resource management and a multiple-access radio resource management. In these hierarchical schemes, local resource managers of the various wireless technologies provide interaction with the central entity in order to perform a joint optimization of available resources. Parallel to the evolution of cellular data systems, the evolution of WLAN, i.e., IEEE 802.11 networks started. The 802.11 systems handovers between different AP from a common domain are based on a decision by the user. Large 802.11 networks show the emergence of problems in resource management in areas with a dense distribution of AP. The concept of unified wireless network architecture argues that the centralized management of data resources in 802.11 networks is necessary to achieve true scalability. WiFi networks were never intended for use as broadband wireless networks. On the other side, WiMAX is also IEEE standard, but with more control implemented in base station. The base station is responsible for fair distribution of available resources among users through the implementation of a centralized system for scheduling. The WiMAX handover functionalities are supported by each base station and the handover is done with their assistance, namely each serving base station helps the user to find their target base station from the list of candidates for base stations to switch.

Taking into consideration the parallel existence of various wireless technologies, the concept of the emergence of heterogeneous networks is not new. The main concept in heterogeneous wireless networks is "Always best connected" (always associated with the best quality), aiming at client terminals. The concept is widely used in various researches. This approach leads to the emergence of vertical handover between different radio access technologies [5]. Based on different optimization techniques such as balancing the load on the network and/or maximizing battery life of the user terminal, it can change the access technology from one to another on a periodic basis or triggered by a given event. In order to perform controlled client assistance at the stage of vertical handover, IEEE created the 802.21 standard referred to as mechanism for exchanging messages between the client and the corresponding base station or AP below the network layer (which is IP in all these cases).

Reviewing the concept of heterogeneous networks inevitably raises the question of inter-working among the radio access technologies in a newly designed system, which will not demand changes in the RATs, but only introduction of control functionalities in the core networks. In terms of the user or user applications, heterogeneous system or a heterogeneous network is considered as a unified network [6] and represents single segment which will place the connection with the application servers in and out of operator’s network. In order to meet the relevant requirements of the user applications generally two possible models are considered for interoperability between building blocks of radio access technologies within the heterogeneous system. The first one refers to a centralized operator access, while the second one defines the Internet model of interoperability. The first model introduces a certain level of integration between the radio access technologies, through which the mobile access terminal realizes handovers. In this direction different analysis have been made and different standards have been developed that should define the levels of architecture connectivity for realizing vertical handover between different access technologies involved in the construction of heterogeneous domain, [5], [6]. The introduction of this model implies to an interoperability protocol of lower levels of communication in the field of radio access. The second model is called the Internet model, formulated and described in detail in this book and refers to providing continuity of customer service in case of independent radio access technologies available to the mobile terminal by connecting on the network level, [7]. In this case, interoperability between network technologies is done on the upper (network) protocol levels, i.e. at a level that is common to all access technologies for communication between user applications with the appropriate application servers.

The ultimate goal of both models for interoperability is the same and that is providing a transparent transfer of user information between client applications and related application servers without impact on the diversity of access technologies in the communication process and providing continuity of user sessions in the communication process. The main difference between the two models is the way of providing interoperability. Besides this difference, the way vertical handovers between access technologies and the conditions or circumstances which trigger handovers are handled is very important. The first method provides an integrated architecture of radio access technologies that builds heterogeneous network, and as such is applicable in cooperative networks or in networks where the radio access technologies are owned by the same operator or cooperating operators. In such networks, there are strictly defined rules for vertical handovers, mainly dictated by conditions in the radio access networks, or by the operator's preference, while user preferences are taken into cooperative architectures. The second method is more general and relates to interoperability regardless of the user's operators, which provides access technology for the user equipment. Generally speaking, vertical handover in these methods is accomplished as a result of the conditions under which user applications see main qualitative parameters of service or experience to the user [7]. The tendency of introducing heterogeneity in future wireless radio systems entails the implementation of different radio interfaces in the new terminals. Each radio access technology has its own radio resource management and they are well engineered for maximum utilization of available resources. Radio access technologies can ensure achievement of customer service in the access part. In most of the radio access technologies which have been made, the system makes adaptation of appropriate resources allocated according to the nature of the services. Considering these characteristics of radio access technologies and taking into account the heterogeneity of future wireless networks and the need for the user to have the best possible quality of its services for a satisfactory price, the need for parallel use of the variety of access technologies emerges, in order to realize the user requirements, [4]. The heterogeneity of these networks allows the user terminal to select the radio access technologies depending on given preferences, [9]. This choice provides better conditions for user applications.

Design of 5G Network Architecture

Figure 1 illustrates the system model of a proposed design of network architecture for 5G mobile systems, which is all-IP based model for wireless and mobile networks interoperability. The system consists of a user terminal (which has a crucial role in the new architecture) and a number of independent, autonomous radio access technologies. Within each of the terminals, each of the radio access technologies is seen as the IP link to the outside Internet world. However, there should be different radio interface for each Radio Access Technology (RAT) in the mobile terminal. For example, if we want to have access to four different RATs, we need to have four different access-specific interfaces in the mobile terminal, and we need to have all of them active at the same time, for this architecture to be functional.

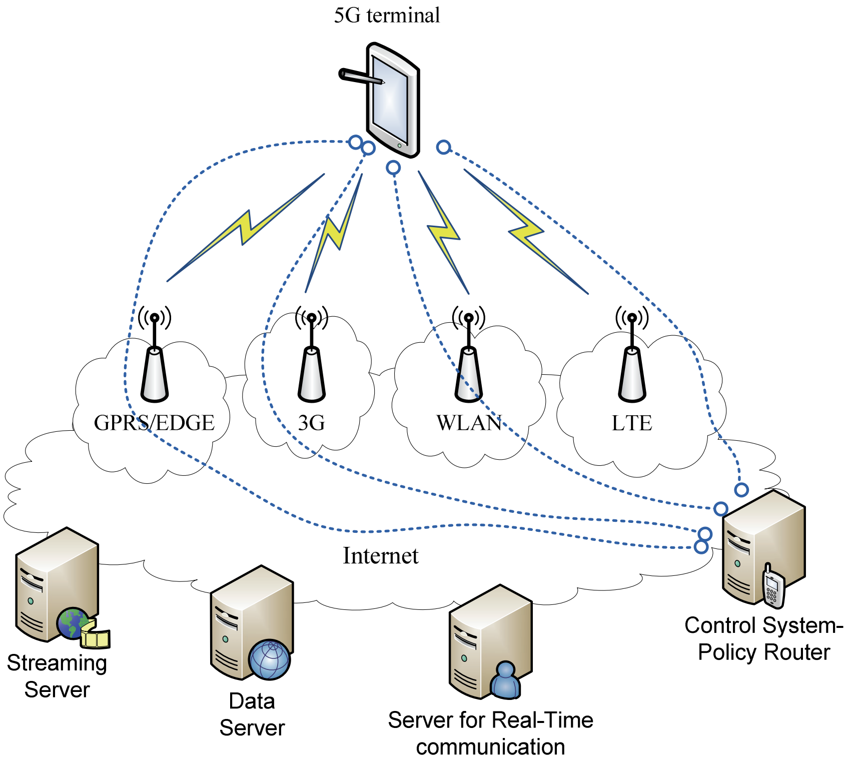


Figure 1. Functional architecture for 5G mobile networks

The first two OSI levels (data-link level and physical level) are defining the radio access technologies through which access is provided to the Internet with more or less QoS support mechanisms, which is further dependent on the access technology (e.g., 3G and WiMAX have explicit QoS support, while WLAN doesn't). Furthermore, over the OSI-1 and OSI-2 layers is the network layer, and this layer is IP (Internet Protocol) in today's communication world, either IPv4 or IPv6, regardless of the radio access technology. The purpose of IP is to ensure enough control data (in IP header) for proper routing of IP packets belonging to a certain application connections - sessions between client applications and servers somewhere on the Internet. Routing of packets should be carried out in accordance with established policies of the user. Application connections are realized between clients and servers in the Internet via sockets.

Internet sockets are endpoints for data communication flows. Each socket of the web is a unified and a unique combination of local IP address and appropriate local transport communications port, target IP address and target appropriate communication port, and type of transport protocol. Having this into consideration, the establishment of end-to-end communication between the client and the server using the Internet protocol is necessary to raise the appropriate Internet socket uniquely determined by the application of the client and the server. This means that in case of interoperability between heterogeneous networks and in the vertical handover between the respective radio technologies, the local IP address and destination IP address should be fixed and unchanged. Fixing of these two parameters should ensure handover transparency to the end-to-end Internet connection, when there is a mobile user at least on one end of such connection. In order to preserve the proper layout of the packets and to reduce or prevent packets losses, routing to the target destination and vice versa should be unique and by using the same path. Each radio access technology that is available to the user in achieving connectivity with the relevant radio access is presented with appropriate IP interface. Each IP interface in the terminal is characterized by its IP address and netmask and parameters associated with the routing of IP packets across the network. In regular inter-system handover the change of access technology (i.e., vertical handover) would mean changing the local IP address. Consequently, change of any of the parameters of the socket means changing the socket, that is, closing the socket and opening a new one. In other words, ending the connection and starting e new one. This approach is not-flexible, and it is based on current Internet communication.

In order to solve this deficiency we propose a new level that will provide abstraction of network access technologies to higher layers of the protocol stack. This layer is crucial in the new architecture.

To enable the functions of the applied transparency and control or direct routing of packets through the most appropriate radio access technology, in the proposed architecture we introduce a control system in the functional architecture of the networks, which works in complete coordination with the user terminal and provides a network abstraction functions and routing of packets based on defined policies. At the same time this control system is an essential element through which the quality of service for each transmission technology can be determined. The system is on the Internet side of the proposed architecture, and as such represents an ideal system to test the qualitative characteristics of the access technologies, as well as to obtain a realistic image of the quality that can be expected from user applications towards a given server in Internet (or peer). Protocol setup of the new levels within the existing protocol stack, which form the proposed architecture, is presented in Figure 2.

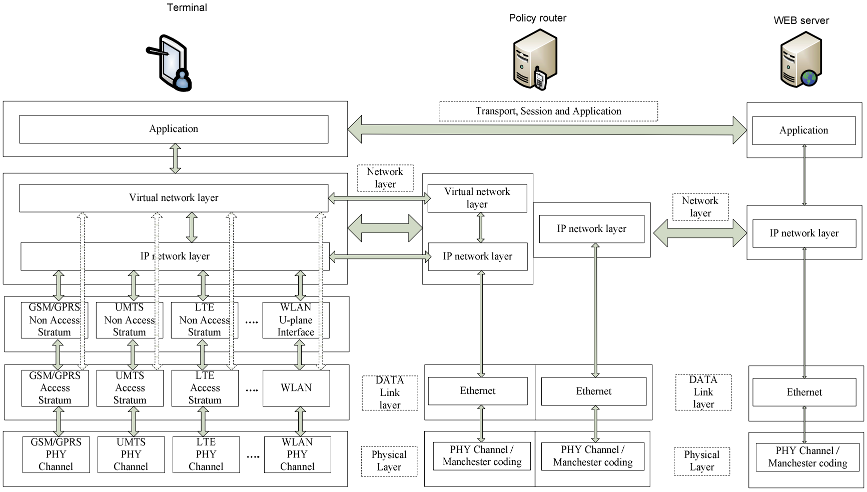


Figure 2. Protocol layout for the elements of the proposed architecture

The network abstraction level would be provided by creating IP tunnels over IP interfaces obtained by connection to the terminal via the access technologies available to the terminal (i.e., mobile user). In fact, the tunnels would be established between the user terminal and control system, called Policy Router, which performs routing based on given policies. In this way the client side will create an appropriate number of tunnels connected to the number of radio access technologies, and the client will only set a local IP address which will be basis for socket forming for Internet communication of client applications with Internet servers. The way IP packets are routed through tunnels, or choosing the right tunnel, would be determined by policies whose rules will be exchanged via the virtual network layer protocol. This way we achieve the required abstraction of the network to the client applications at the mobile terminal. The process of establishing a tunnel to the Policy Router for policy based routing, is carried out immediately after the establishment of IP connectivity across the radio access technology, and it is initiated from the mobile terminal Virtual Network-level Protocol. Establishing tunnel connections as well as maintaining them represents basic functionality of the virtual network level (or network level of abstraction).

Description of use-cases in the proposed network architecture

Heterogeneity of wireless networks enables the user terminal to select access technologies depending on their preferences. This choice provides better conditions for user applications. The processes of achieving connectivity in new environments are strongly associated with the application process. Namely, the need of the user application to establish communication with an application server usually ends by initiating a connection through the network level, i.e., network access to resources by the user terminal.

Considering that the functions of the virtual network layer in the proposed new architecture include many functions related to connectivity, security and continuity of the application sessions initiated by the user, the virtual network layer is divided logically into several cooperative software modules which perform different functionalities. A block-diagram of the software modules in the virtual network layer is illustrated in Figure 3.

There are several differences between client and server functions to a virtual network layer. On the client side there are five software modules:

- RAT-CCSM (Radio Access Technology - Connection Control Software Module);
- MQPBR (Mobile Quality Policy Based Router);
- SPME (Security and Policy Management Entity);
- ITHC (Inter Tunnel Handover Control); and
- QoS / QoE CM (QoS and QoE Control Manager).

On the other side, the Policy Router includes four software modules as follows:

- MCCSM (Media Connection Control Software Module);

- CQPBR (Central Quality Policy Based Router);
- SPME (Security and Policy Management Entity);
- CPH (Client Profile Handler); and
- QoS / QoE CM (QoS and QoE Control Manager).

Each software module has specific position within the global architecture to provide ultimate functionality and interoperability in 5G heterogeneous systems. The functionality between software modules is provided through precisely defined interfaces to other modules and with appropriate links between peer protocol modules on both sides of the architecture.

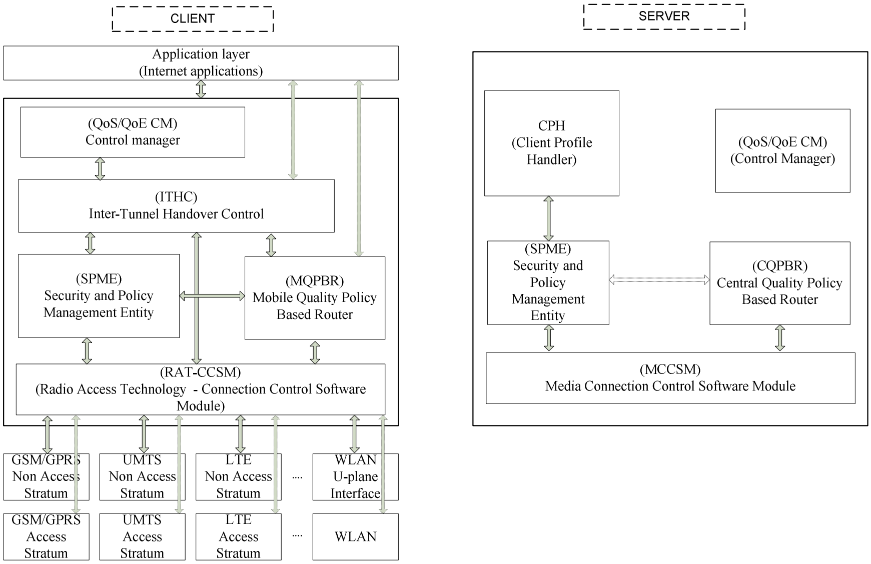


Figure 3. Software diagram of the proposed virtual network layer

II.1. Functionality of the virtual network layer in the proposed architecture

As it can be seen from the diagram in Figure 3, there are four common cooperative/coordination modules on different sides (client and server) that are interconnected. Hence, we may distinguish among four basic functionalities carried out by the virtual network layer.

The first basic functionality of the virtual network layer is to provide a network abstraction. This functionality is related to the cooperative working together of the RAT-CCSM and MCCSM software modules that are designed to make masking of the IP level seen by each radio access technology. Besides this basic functionality, RAT-CCSM module at the client side is using API interfaces for access to the appropriate software modules from the lower levels of radio access technologies in order to provide additional information. This link is a way through which it receives information for improving connectivity of individual access technologies (e.g., generated PDP context with specified IP address, connection established with a given AP in WLAN networks and corresponding IP addresses, etc.) and the level of received signal of the corresponding radio access technology. This way, the software module has continuous information for the network and radio conditions in each radio access technology. Tunnels are formed between RAT-CCSM on the client side and MCCSM module in the Policy Router. RAT-CCSM module starts a process to establish a tunnel between the mobile client and the Policy Router (in particular, with the MCCSM module). The tunnel is formed through the established IP connectivity of the particular radio access technology. Source IP address of the tunnel is the IP address obtained by establishing IP connectivity via the given access technology, while destination IP address of the tunnel (from the uplink direction on the mobile side) is the loopback address of the software module of MCCSM Policy Router. RAT-CCSM software module performs continuous monitoring of the status of each radio access technology in terms of radio parameters (signal received level) and of IP connectivity through the same network. The obtained information about the radio access technologies is forwarded to another associated software module whose primary function is handover management in the transmission of data between the established IP tunnels (the ITHC software module). The second link of this module refers to the routing module, where routing is based on policies determined on the offered Quality of Service. Their interaction results in defining appropriate tunnel interfaces (corresponding to the tunnels created by the radio access technologies) within the routing table. The process of establishing the tunnel procedure begins with authentication and authorization between the mobile client and the Policy Router, so the software module has a direct connection with SPME module for management of security mechanisms.

The second function is related to routing policies based on the determined Quality of Service offered by access technologies. This functionality is accomplished by cooperative working between MQPBR and CQPBR software components on the client and server sides. The mutual cooperation between these two modules is realized through the appropriate routing control protocol developed specifically for this purpose. Its goal is to provide proper prioritization of routes or routing rules via the tunnel interfaces within routing matrix / table. The changes are initiated and controlled by MQPBR client module in cooperation with the ITHC module. At the same time, the MQPBR software module on the client side handles the client IP address which is obtained in the phase of the authentication and authorization by the software module for that purpose - SPME. The actual determination of the client mobile IP address will be marked with McIP, which is an IP address of the client in the heterogeneous network generated by SPME software module and subsequently it is given to the MQPBR client module. The communication of the upper protocol levels, such as the transport, session and application levels, is maintained via McIP address of the user, which is seen as single permanent IP network address. The main feature of this software module, in comparison with other routing software components, is its ability to perform coordinated routing between the two software modules depending on the application that is initiated by the client. This would mean that routing table of this module expands and takes the form of three-dimensional routing matrix where for each initiated user application priority for routing over the tunneling interface is defined.

The third function is associated with managing security procedures or security mechanisms and policies applied to users. RAT-CCSM module triggers corresponding module on the client side (SPME) in order to carry out proper user authentication and authorization in the process of creating a tunnel through the appropriate technology. This process is accomplished through any “free” IP address obtained from a radio access technology towards a defined IP address of the server on the other side. In this case RAT-CCSM transparently forwards these packets directly to the network interfaces of the radio access technologies. After receiving the result of a process of authentication and authorization RAT-CCSM and MCCSM begin the process of establishment of an IP tunnel or reject the request depending on the result of the authentication process. On the client side, the user terminal contains all the information in a local storage (in the mobile terminal) within the security software module, while the Policy Router stores the information for the mobile clients in an additional software module, referred to as CPH, which can be part of the same Policy Router (however, it is not mandatory). All information for each user of this architecture, the authentication parameters and policies, are stored in this database - CPH. All policies and user parameters that describe a customer, which are obtained from other systems and stored in CPH module, are made available to RAT-CCSM module as well as MQPBR and CQPBR modules and the IHTC module. The RAT-CCSM module establishes a tunnel; the defined McIP address is assigned to MQPBR and CQPBR modules, while other policies are assigned to IHTC which are contained in the CHP that should help in the process of handover decisions.

The fourth functionality is associated with the management mechanisms for measuring of the parameters that define the Quality of Service and Experience in terms of user applications. This functionality is accomplished by cooperative working between the QoS/QoE module on the client side and QoS/QoE module on the server side. The purpose of this module in the mobile terminal (the client side) is to continuously measure the basic qualitative parameters of radio access technologies. Thus, the measured parameters give a realistic picture of the Quality of Service that can be expected from the radio access technologies, which in fact are on the path between the client and Policy Router. Measurements are carried out individually by each access technology. The results of these measurements are a direct input to the ITHC module for handover decisions between tunnels.

Fifth functionality of the network architecture is dedicated to the user only, and its location within the heterogeneous wireless network. This functionality is intended to ensure continuity of customer service while taking into account the qualitative requirements of the applications, the user, and the network, in a form of predefined policies or gained knowledge from the user services. This module on the user side is represented as ITHC software module and has a direct interaction with other software modules of the virtual network layer. Software Module continually processes data from RAT-CCSM software module (realized tunnels and signal reception level of each access technology). Also, it is directly associated with the QoS/QoE module, from which it receives information about the qualitative characteristics of each radio access technology used by the user. Then, with aim to decide which application will use which available radio access technology, it receives from the SPME the user policies as well as preferences of the user and the operator (that is the one that provides the functionalities of Policy Router). If there is a need of change of the access technology for an ongoing session, this module is required to initiate the process of handover between tunnels connected with the relevant access technologies. The criteria under which it will begin the procedure of handover are part of the software module and its internal logic. The change of priorities for the routes for each application is performed by the module responsible for policy-based routing, i.e., the Policy Router on the network side.

In the following part we present more important use-cases of the proposed architecture in this section.

II.2. Establishing a tunnel

The process of establishing the tunnel begins with the first phase, a phase of continuous monitoring of the Radio Access Technologies (RATs), performed by RAT-CCSM module. At this stage of the process, it is cyclically repeated interrogation of the RATs through direct interaction with their network level. This activity is covered in steps 1 and 2 of the diagram in Figure 4.

In the second phase, when the mobile terminal detects established IP connectivity to the network layer of a RAT, represented by step 3, begins the process of authentication and authorization of the client and relevant RAT (step 4). In this step there is interaction with the module for managing security mechanisms and policies SPME. Failure of authentication results in error that should be presented to the user (steps 5 and 6).

The third phase of the process of establishing the tunnel starts with a successful authentication and authorization of the client-side using the Policy Router (step 7). This is a trigger for the start of step 8, which is initiation of the procedures for establishing a tunnel between client and server interacting with MCCSM module on the side of the Policy Router. Failure of this procedure (step 9), initiates the mechanism of recurrence of the process that repeats itself via steps 8, 9, and 10, retrying until it exceeds the permitted number of attempts defined in the algorithm.

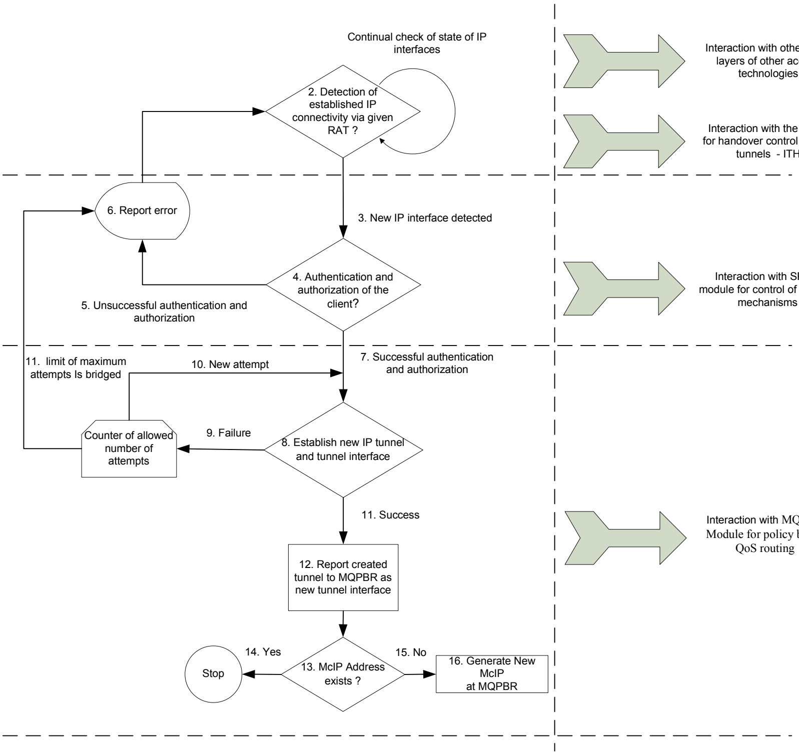

Continual check of state of IP interfaces
2. Detection of established IP connectivity via given RAT ?
Interaction with other radio layers of other access technologies
Interaction with the module for handover control between tunnels - ITHC
6. Report error
3. New IP interface detected
4. Authentication and authorization of the client?
5. Unsuccessful authentication and authorization
Interaction with SPME module for control of security mechanisms
11. limit of maximum attempts Is bridged
10. New attempt
7. Successful authentication and authorization
Counter of allowed number of attempts
9. Failure
8. Establish new IP tunnel and tunnel interface
11. Success
Interaction with MQPBR Module for policy based QoS routing
12. Report created tunnel to MQPBR as new tunnel interface
14. Yes
15. No
Stop
13. McIP Address exists ?
16. Generate New McIP at MQPBR

Figure 4. Block diagram for establishment of Internet tunnel for a given radio access technology

This phase ends with the successful creation of the tunnel and the corresponding tunnel interface. If this represents the first created tunnel, i.e., there is no McIP address assigned to the user, then the IP address obtained during authentication and authorization of the user at the Policy Router is the one that is given to the MQPBR module with aim to be used as source address for communication initiated by the user.

II.3. Establishment of customer service

The process of establishing of customer service starts with the first phase, a phase of the user initiating the connection. The user in this process is accessing to its own applications and starts the interaction with the appropriate application server. This activity is covered in steps 1 and 2 of the diagram shown in Figure 5. At this stage of the process the application level is in direct interaction with the virtual network layer of the proposed architecture, specifically with its module for policy-based routing packets and Quality of Service - MQPBR.

The establishment of customer service from the mobile terminal is outlined in diagram shown in Figure 5.

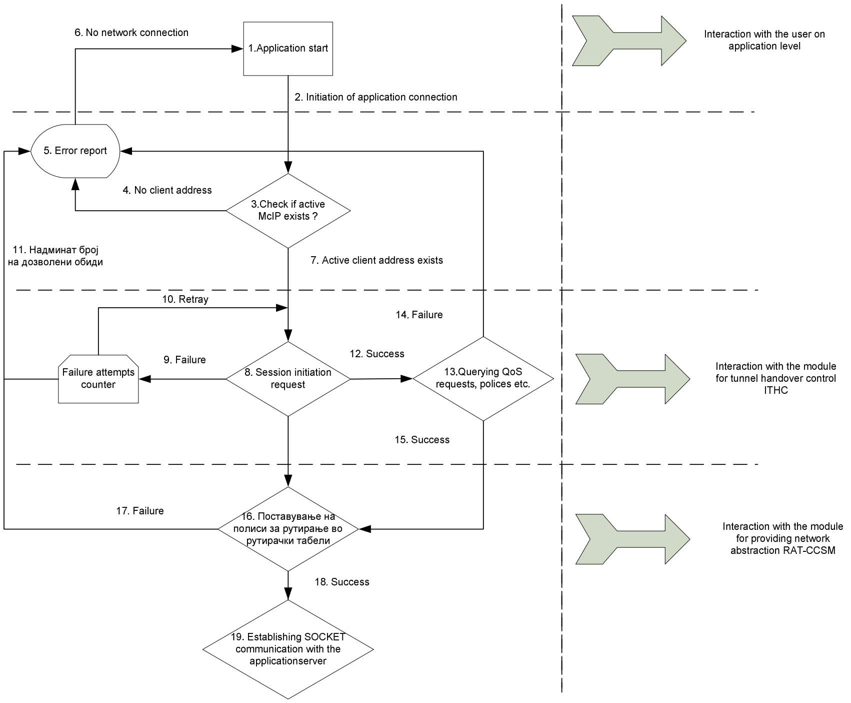

6. No network connection
1.Application start
2. Initiation of application connection
Interaction with the user on application level
5. Error report
4. No client address
3.Check if active McIP exists ?
11. Надминат број на дозволени обиди
7. Active client address exists
10. Retray
14. Failure
12. Success
9. Failure
Failure attempts counter
8. Session initiation request
13.Querying QoS requests, polices etc.
Interaction with the module for tunnel handover control ITHC
15. Success
17. Failure
16. Поставување на полиси за рутирање во рутирачки табели
Interaction with the module for providing network abstraction RAT-CCSM
18. Success
19. Establishing SOCKET communication with the applicationserver

Figure 5. Block diagram for establishing customer service

II.4. Process for changing Radio Access Technology - tunnel handover

The process of changing radio access technology begins with the first phase, a phase of collecting information as input for the handover algorithm. This activity is covered in steps 1, 2, 3, 4 and 5 of the diagram in Figure 6 that is performed in direct interaction with other components of the virtual network layer (RAT-CCSM, QoS / QoE CM, SPME), as well as using user preferences. At this stage of the process all the above components are in direct interaction with the ITHC. The second phase of the process presents an analysis of the input parameters through an intelligent algorithm in step 6, which main function is to manage handover. This procedure is repeated for each user session. The algorithm used for decision making mechanism is shown in Figure 7.

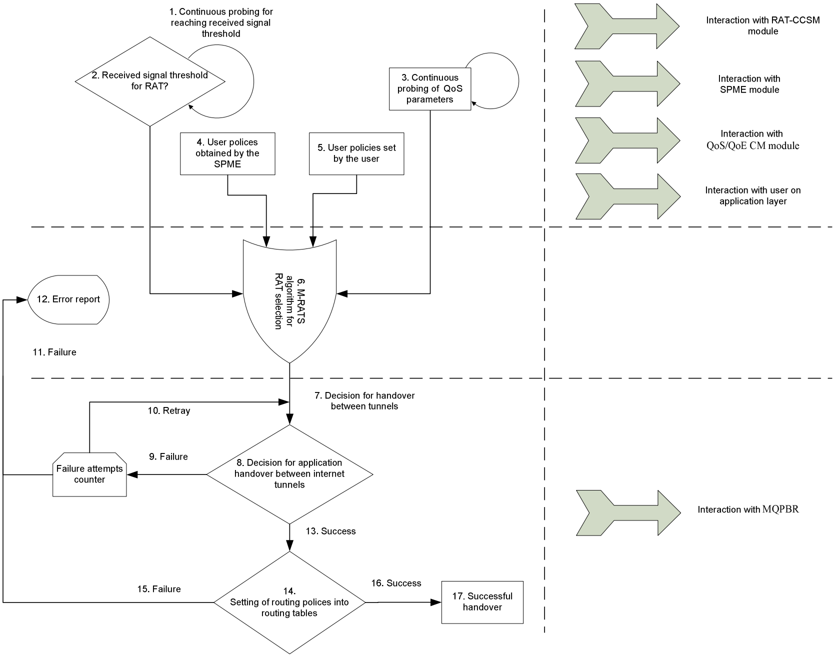

1. Continuous probing for reaching received signal threshold
2. Received signal threshold for RAT?
3. Continuous probing of QoS parameters
4. User polices obtained by the SPME
5. User policies set by the user
6. M-RATS algorithm for RAT selection
7. Decision for handover between tunnels
8. Decision for application handover between internet tunnels
9. Failure
10. Retray
11. Failure
12. Error report
Failure attempts counter
13. Success
14. Setting of routing polices into routing tables
15. Failure
16. Success
17. Successful handover
Interaction with RAT-CCSM module
Interaction with SPME module
Interaction with QoS/QoE CM module
Interaction with user on application layer
Interaction with MQPBR

Figure 6. Block diagram for changing RAT – handover between IP tunnels

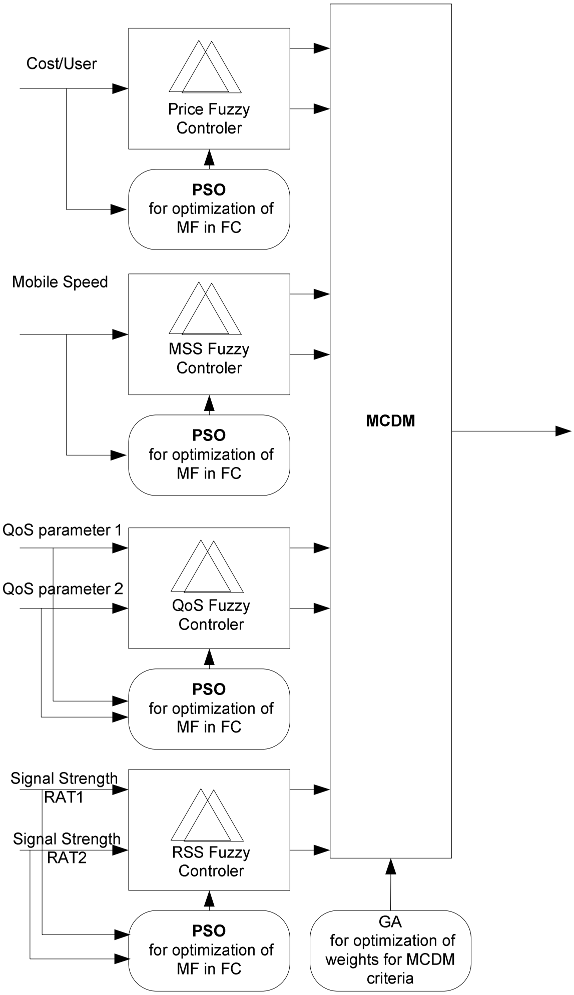

Cost/User
Price Fuzzy Controler
PSO for optimization of MF in FC
Mobile Speed
MSS Fuzzy Controler
PSO for optimization of MF in FC
QoS parameter 1
QoS parameter 2
QoS Fuzzy Controler
PSO for optimization of MF in FC
Signal Strength RAT1
Signal Strength RAT2
RSS Fuzzy Controler
PSO for optimization of MF in FC
MCDM
GA for optimization of weights for MCDM criteria

Figure 7. User software agent for RAT selection by the mobile terminal

II.5. Authentication and authorization of access technology with policy exchange.

The process of authentication, authorization and exchange of policies when connecting to certain access technology begins with the first phase, which is phase of initiation request received in interaction with network abstraction module RAT-CCSM. This activity is covered by steps 1 and 2 of the diagram shown in Figure 8.

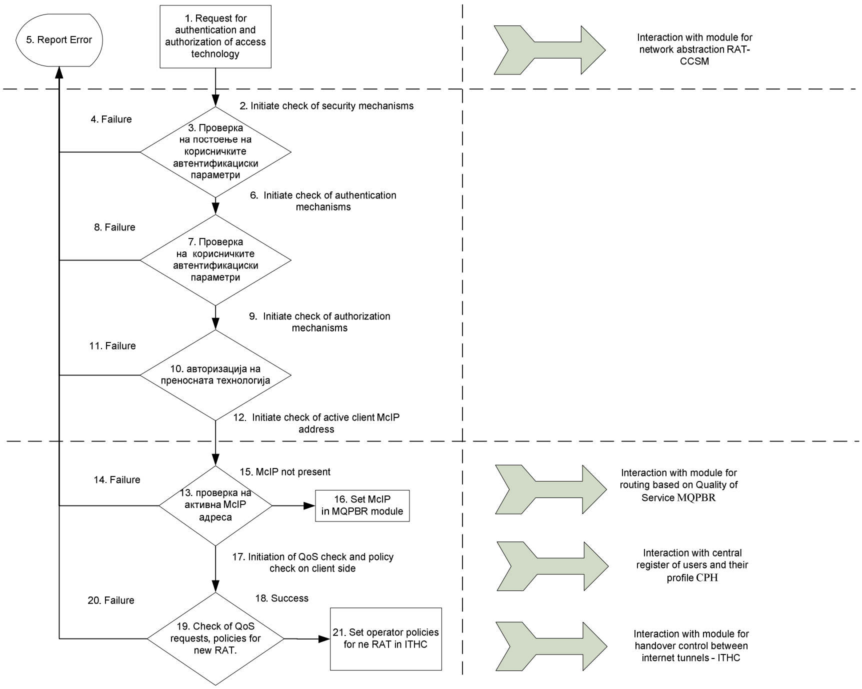

5. Report Error
1. Request for authentication and authorization of access technology
Interaction with module for network abstraction RAT-CCSM
2. Initiate check of security mechanisms
4. Failure
3. Проверка на постоење на корисничките автентификациски параметри
6. Initiate check of authentication mechanisms
8. Failure
7. Проверка на корисничките автентификациски параметри
9. Initiate check of authorization mechanisms
11. Failure
10. авторизација на преносната технологија
12. Initiate check of active client McIP address
14. Failure
15. McIP not present
13. проверка на активна McIP адреса
16. Set McIP in MQPBR module
Interaction with module for routing based on Quality of Service MQPBR
17. Initiation of QoS check and policy check on client side
Interaction with central register of users and their profile CPH
20. Failure
18. Success
19. Check of QoS requests, policies for new RAT.
21. Set operator policies for ne RAT in ITHC
Interaction with module for handover control between internet tunnels - ITHC

Figure 8. Block diagram for authentication and authorization of access technology and policy exchange

In the second stage, a mechanism for authentication and authorization is activated, which is essentially under control of the module for management of the security mechanisms and client policies (SPME) placed on client side, and the corresponding module on the policy-router side. SPME modules are in close cooperation with RAT-CCSM and MCCSM levels because during the process the tunnel establishment through certain radio access technology, RAT-CCSM module triggers appropriate SPME module on client-side in order to perform proper authentication and authorization of user for creation of a transport tunnel through appropriate technology.

This process is accomplished through any free IP address obtained from a radio access technology to the IP address defined by the server. The first step in achieving this goal is to check the user authentication parameters by the client (step 2). This involves verifying the authentication resources that for the proposed architecture based on user certificates involves checking the existence of the user certificate. The validation of user certificate must be located in the part of the software module designated as MIM (Mobile Identity Management). In this step (step 3) validity check of the certificate, its private and public keys and the validity of the issuer and its purpose is conducted. In case of wrong parameters or error during these checks, an error is raised and sent towards initiating level, in this case the RAT-CCSM, in order to stop the process of establishing a tunnel, step 4 and 5. Once the validity of user parameters are determined, the flow continues through the process of authentication and authorization by SPME modules. Communication between two modules is based on PANA protocol where authentication of user preferences is performed through EAP-TLS or EAP-SAML. In this case, function of EAP client is performed by user module, while function of authenticator is conducted by SPME module in policy-router where CPH module will be presented as DIAMETER server as it represents the central registry of user data and their profiles. Procedures for authentication and authorization of access technology are presented as step 7 and step 10. In the event of a failure, error is reported as presented in step 4. After the authentication process ends, the flow continues by checking the existence of client address McIP by MQPBR entity. In this way, in step 13, check of network layer is conducted, that determines whether this is the first tunnel through which connection towards policy-router is realized. Completing the process of authentication and authorization initiates the process of exchange of policies and user parameters (McIP address client). This process is performed in the second phase, through the appropriate protocol. If there is a client address process, it continues to the screening phase of QoS requirements and policies within the ITHC module for concrete tunnel, in step 21. The initiated process ends with this action.

II.6. Termination of user service or user session

The process of terminating a customer service begins with the first phase, which is the stage where the user initiates closing of the application or customer service. The user in this process accesses its applications and starts process of interaction with the appropriate application server in order to terminate the established service. This includes start up procedures for properly closing the service between the user and the server application level. This activity is covered by steps 1 and 2 of the diagram shown in Figure 9. At this stage of the process, the applicative level is in direct interaction with the virtual network layer of the proposed architecture, specifically with its policy-based routing module and Quality of Service - MQPBR. In essence, this module has complete control over this process. Description of the process is shown in Figure 9.

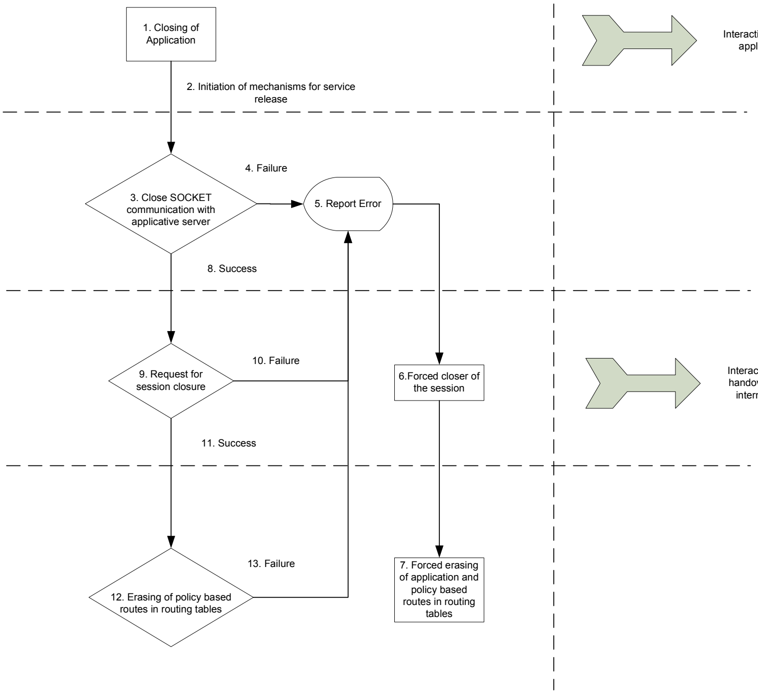

1. Closing of Application
2. Initiation of mechanisms for service release
Interaction with user on application layer
3. Close SOCKET communication with applicative server
4. Failure
5. Report Error
8. Success
9. Request for session closure
10. Failure
6.Forced closer of the session
Interaction with module for handover control between internet tunnels - ITHC
11. Success
12. Erasing of policy based routes in routing tables
13. Failure
7. Forced erasing of application and policy based routes in routing tables

Figure 9. Block diagram for termination of a customer service

In the second phase of the process, communication mechanism for terminating of the user service (step 3) is triggered which results in termination of transport connection with application server. If the cancellation is successful, the process continues with session termination created in ITHC module in order to ensure continuity of customer service and in the same time to maintain the required performance of the service (steps 8 and 9). At this stage direct interaction with ITHC module is achieved in order to successfully terminate user session. The success of this step allows transfer to the third and final stage where the process of deleting the application and policy based routes in the MQPBR software module (step 12) starts. If an unsuccessful attempt appears in any of the above steps 3, 9 and 12, the adequate working level reports an error and starts with the mechanism of involuntary termination of customer service that leads to the forced closure of the user session in ITHC module (step 6) and forced deletion of applicative and policy based routes in MQPBR (step 7). After completion of the process of termination of a customer service in software modules of the virtual network layer of the proposed architecture, all information of its existence are deleted, and all links that are created and set for its maintenance are broken.

Innovative design of network layer for transparent connectivity of higher network layers over heterogeneous networks.

In the review of the basic design of the proposed architecture we have explained that the essential part in the process of providing internet interoperability of heterogeneous networks is introduction of the network abstraction between the client heterogeneous networks and the policy-router. The network abstraction level would ensure the creation of IP tunnels over IP interfaces acquired by terminal connection over different access technologies. Next section provides an overview of different types of technologies for establishing of Internet tunnel between two network entities. The purpose of this analysis is to choose the best technology that would satisfy the requirements of next-generation mobile networks, providing maximum security and flexibility at the same time.

II.7. Types of tunneling protocols and their use

The idea behind the Internet tunnel is to create a tunnel connection that will simulate a private link via shared network using encryption and tunneling techniques. The creation of the network, which is based exclusively on IP protocol stack, increases the number of independent network segments through which IP packets are transported between source and destination. To achieve direct (point-to-point) communication, tunneling of IP packet is introduced. Tunneling is a method of using infrastructure interworking between network segments to transfer data from one network through network infrastructure of another network. Rather than sending the packet as it is produced by the source, tunneling protocol encapsulate the packet with additional header, as shown in Figure 10, and forwards it further to its destination [25]. This way encapsulated packets are routed between the two ends of the tunnel (tunnel interfaces) through the existing internet network. The logical path through which encapsulated packets travel over the transport network is called "tunnel". Once encapsulated packets are received on the other side of the tunnel, they are decapsulated in order to obtain the original packets. Tunneling of the packets encompasses the entire process of encapsulating, transfer and decapsulating of the packets in communication between two points.

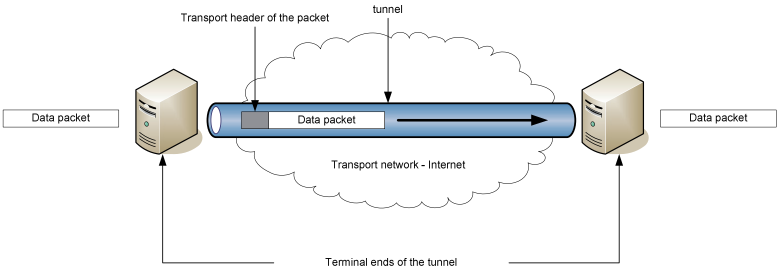

Transport header of the packet
tunnel
Data packet
Data packet
Data packet
Transport network - Internet
Terminal ends of the tunnel

Figure 10. Transport technique for tunneling of data packets

Tunneling of data packets can be achieved over any of the existing data networks, but in the real world usually the most considered case of tunneling of packets is over global Internet network, or network based on IP protocol. [26] This approach, tunneling through the Internet to provide interworking operation in heterogeneous networks, is applied in the case of the proposed architecture.

To create a tunnel between two points, the tunnel client and the tunnel server must use the same tunneling protocol. Tunneling can take place at different levels according to the OSI protocol stack, and in the context of the new architecture we consider tunneling on 2nd and 3rd level by OSI.

Second level tunneling protocols correspond to the OSI data link level and they use data frames as basic units for data exchange. Protocols that operate on this OSI layer are: PPTP (Point to Point Tunneling Protocol) and L2TP (Layer 2

Tunneling Protocol), which encapsulate the packets in PPP (Point to Point Protocol) and send them over the transport network.

Tunnel protocols on third level correspond to network layer according to the OSI and use data packets as basic units for data exchange. Protocols at this level perform encapsulation of packets in new packets by adding additional headers before they are forwarded through the transmission network. Tunnel protocols that operate on this layer are protocols like: IP in IP, GRE (Generic Routing Encapsulation), GTP (GPRS Tunneling Protocol), IPSec, e.t.c.

In order to make the proper selection of tunnel protocol that will provide network abstraction according to the proposed architecture we will give a basic overview of the features of each of the protocols that can be used for this purpose.

PPTP (RFC2637) protocol is designed to provide secure transmission of data from a remote client to a private corporate server in a way that it creates a virtual private network over the underlying IP transport network [26]. Generally PPTP protocol realizes regular PPP (Point to Point Protocol) session through GRE protocol. Besides this there is a second session via TCP session that is used for initiation and management of data, transport and GRE session. PPTP protocol as authentication mechanism for establishing of PPP uses some of the standard authentication mechanisms EAP-TLS or MSCHAP-v2.

L2TP (RFC2661, version 3 RFC3931) [27] combines the best features of PPTP, d [28] and (L2F - Layer 2 Forwarding) protocol [29]. L2TP uses UDP protocol for transmission of data packets and is transparent to the protocols of the upper levels. Tunnel is established between the LAC (L2TP Access Concentrator) and LNS (L2TP Network Server), and once established, the network traffic through the same tunnel is bidirectional.

IP in IP tunneling protocol (RFC 1853, RFC 2003) is the basic type of L3 tunneling in IP networks. Source data IP packet, in this tunneling protocol that needs to be transmitted between the two ends of a tunnel is encapsulated in another IP packet. The underlining packet which transfers the data IP packet, is marked as transport IP packet. Internal user packet header does not change by encapsulation except for TTL field which is decremented by 1 with encapsulation if the destination user packet is through the established tunnel. Decapsulation does not change the TTL (Time Life) of extracted packet. In case of user packet with TTL = 0, encapsulation should not be performed, similar, if after decapsulating of the packet, TTL parameter of the internal packet data is equal to 0, the packet should be removed. Other areas that are of importance for the treatment of inner packet as "TOS (Type of Service)" and "Don't Fragment" are copied from the internal to the external packet. If the length of the source packet is greater than transmission MTU (Maximum Transmission Unit) of the transport path over which the tunnel is formed then fragmentation of the user packet is performed by the tunnel protocol. Final reassemble of the original packet is performed later during decapsulation process. In this protocol there is no additional tunnel management beside the usual ICMP mechanisms. This tunnel is not working between two points that are located behind a NAT (Network Address Translation) gateway. According to the specification in RFC 2003, IP in IP tunneling protocol does not define appropriate authentication mechanism. However certain header authentication can be done between the original inner and outer tunnel transport header packet.

GRE tunnel protocol is designed for encapsulation of an arbitrary protocol over any other arbitrary protocol. GRE enables creation of a tunnel through a network protocol which hides the content of transmitting protocol through the tunnel. Frequently, the implementation of the GRE protocol is highly related to the IP network protocol. Example of the concrete implementation of the GRE protocol is presented in Figure 11.

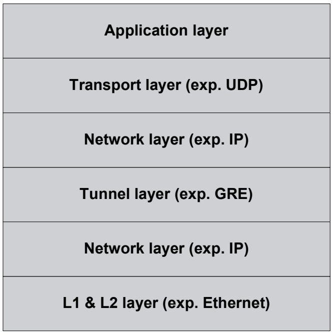


Figure 11. Implementation of GRE tunneling protocol for tunneling of IP protocol over IP network

Generally there are two implementations of the GRE protocol, one based on RFC 1701, and other based on more recent RFC 2784. To some degree RFC 2784 implementations are interoperable with GRE protocol implementations according to RFC 1701, but some of the functionalities in GRE implementation according to RFC 1701 are excluded in RFC 2784. Some of them are recovered by reimplementation in RFC 2890. GRE protocol format according to RFC 1701 and RFC 2890 is shown in Figure 12 and Figure 13.

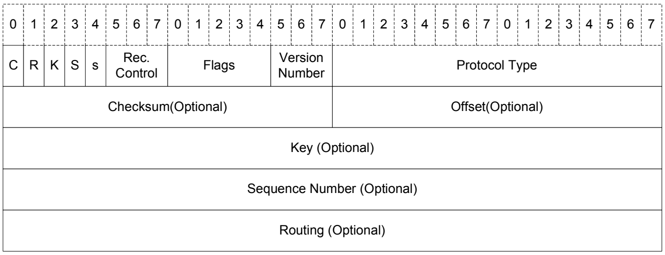


Figure 12. GRE implementation according to RFC 1701.

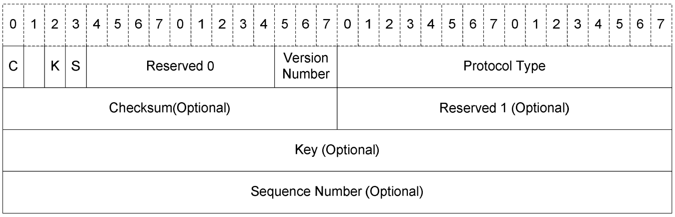


Figure 13. GRE implementation according to RFC 2890.

GRE protocol (RFC 2784) is session-less protocol, where the terminal ends if the tunnel does not monitor the status and availability of the opposite end of the tunnel. Apart from the above mentioned RFC, there are others that describe how to use GRE protocol in different environments. An extension of the basic GRE specification that is of particular importance is the GRE "Key and Sequence Filed" defined in RFC 2890. This protocol does not provide data encryption by itself, but provides possibility for tunneling of packets with a small amount of data in the header.

C field of the GRE packet indicates whether fields “Checksum” and “Reserved1” are present in the packet. In this case the “Checksum” field contains the checksum of the GRE header as well as the data part of the packet. "Reserved1" field, if present in the packet is set to all zeros. If “C” field is not set in the header then the fields “Checksum” and “Reserved1” are not present in the packet. “K” and “S” fields in the GRE header indicate whether fields “Key” and “Sequence” are present in the GRE packet. “Protocol Type” filed contains information about the type of protocol that is found in the data part of the GRE protocol. “Key” field in the header of the GRE protocol is set in order to identify individual data stream in GRE tunnel. GRE protocol does not define a specific way of implementing this field. One possible implementation of this field is within the GRE protocol, specifically within the PIMP-based interfaces. "Sequence" field is defined in order to maintain the sequence of packets within the GRE tunnel. This field is defined by the source in the process of encapsulation, which is used by decapsulation process for proper alignment of the received packets.

GRE protocol is often used together with IPSec protocol from the network layer, in a way that IPSec provides encryption of data that is transmitted through the GRE tunnel. Forwarding of GRE encapsulated packets is not really different from forwarding of normal packets. When a GRE packet is received, first, the receiver must recognize or detect that it is a GRE packet, so once it determines its background, it checks its header and the contents of the packet with the help of information from the header. Afterwards it starts with its decapsulation which removes all protocol fields from the GRE protocol. GRE packet detection is performed using a field which indicates the type of protocol that is inserted into the IP packet. This field has the number 47, which indicates that the data part of the IP packet contains a GRE packet.

GTP protocol can operate via TCP and UDP as transport protocols on fourth level by OSI, and it requires an explicit exchange of signaling messages to establish the tunnel. Initially, in the first version of the GTP protocol - GTPv0, signaling protocol for establishing the tunnel was combined with the tunneling protocol through a common port. This approach was changed in the second version GTPv1, which introduced two protocols, one for the control of the tunnel “GTP-C” and other for tunneling of user data “GTP-U”. Besides these two protocols GTP set of protocols was completed with the third GTP' protocol that is used for charging. Later 3GPP introduced the third version of this protocol GTPv2 [TS 29.274] (AKA evolved GTP or eGTP) developed for implementation in LTE / EPS. GTPv2 also has two protocols, protocol for control of the tunnel “GTPv2-C” and protocol for transmission of user data “GTPv2-U”. GTP tunnel is identified in each network element with appropriate TEID (Tunnel Endpoint Identifier) and the appropriate UDP port. Receiving end sets TEID identifier should be used in communication between the two ends of the tunnel. These values are exchanged between two network entities at both ends of the tunnel by means of exchange of signaling messages through the GTP-C protocol.

IPSec protocol RFC 2401 appears with the later development of the Internet in the early 1990ies by IETF. Over the years it has undergone several upgrades as described in RFC 4301, and its functions are defined in other RFC'a. The use of IPSec protocol in IPv4 networks is optional functionality while in IPv6 networks this functionality is included in the protocol. Sending unprotected IP packet over an IP network makes it vulnerable to various attacks during its routing. These attacks are mainly related to its reliability (it is possible to alter the source and destination IP address of the package), as well as, its authenticity and data integrity (it is possible to change the contents of the package). In order to perform adequate protection against such attacks IPSec protocol was designed to provide appropriate care and services such as: access control, data integrity protection, providing privacy, protection of forwarding packets, and their compression [30], [31], [32].

The term access control stands for a service that provides protection against unauthorized access to a resource such as a server or network. The sender authentication service allows the recipient to verify the sender of data over the network. The service for protection of data integrity allows protection of the session between a client and its server. Through this service it can be determined whether a change of data was made during its transmission across the IP network, but cannot be determined whether the data was copied or duplicated. Detection and rejection of duplicates control is a form of partial sequence integrity of the data, where the receiver can determine if irregularity or a duplication of a given package appears. Ensuring privacy of communication is a service that protects the visibility of the content of the data traffic from external unauthorized parties. The mechanism that ensures privacy in IPSec protocol is mechanism for encryption of data traffic. This involves transforming the contents of the packet by using the encryption algorithm in a way that it becomes incomprehensible to the environment. Besides this full encryption, IPSec protocol also provides privacy and data stream encryption, which implies encryption of only one of the parameters pertaining to the type of service, the source and destination address as well as the length of the packet. To be able to use services provided by IPSec protocol between two systems, both systems need to share certain security parameters, such as: security keys and mechanisms, encryption algorithms, etc. In order to provide proper management of these parameters in IPSec protocol, separate security associations (SA - security associations) are defined. The term security association means a relationship between two entities that define a way of communication between them. Security association denotes unidirectional connection, so to achieve IPSec protection of traffic in both directions it is necessary to create a pair of security associations, one for each direction of communication. Each SA is uniquely determined by three parameters: SPI (Security parameter Index), destination IP address and security protocol (AH or ESP - explained below). Information related to formed security associations are kept in SADB (Security Association Database) where besides security association at each connection point a database for security policies (SDP - Security Policy Database) is defined. SDP keeps the definition of policies that have to be applied on appropriate data traffic to a given destination.

AH protocol provides data integrity, data source verification and protection from duplicated packets. Network source digitally signs the packets in order to perform authentication on the client side. Receiver checks the digital signature of the packets and depending on the outcome of this check either accepts or rejects the packet. If the packet is changed during the transport over the network to its destination path, the digital signature will not match the contents of the packet. The content of the packet is not encrypted, but it is digitally signed.

ESP protocol in addition to the AH protocol provides encryption of data in the packet. This way, it provides complete data privacy in the process of communication between the two ends of the communication. Anyone who would tap the communication between the two ends could capture exchanged packets, but could not decipher their content because the entire contents of the packets is encrypted and further it can be digitally signed with AH.

Management of Security Keys used for encryption is not part of the protocol, but can be provided by user assistance or by using IKE (Internet Key Exchange) protocol [RFC2409] based on a mechanism for the exchange of public keys used for automated key management. IKE protocol (new versions of IKE (SOI) or IKEv2) participates in the process of creating a security links and defines the algorithms and methods for encryption, authentication and data integrity. At the beginning this protocol establishes a shared secret key between two entities that need to establish communication in order to ensure safe and secure communication between them, and then it performs authentication of the communication entities and starts creating security associations (SA). In addition to this technique for automatic distribution of keys other techniques as Kerberos and SKIP can be used.

IPSec protocol provides two operating modes: transport and tunnel mode of operation. Transport mode refers to two entities in direct communication with each other without the presence of traversing entities between them, while tunnel mode refers to protection of entire IP packet during its transport between two entities across the rutting network. AH and ESP protocols within IPSec can be used both in tunnel and transport mode operation. They can work individually or in mutual cooperation in order to meet all safety requirements in IPv4 and IPv6 networks. Description of encapsulation of IP packet using IPSec and IP in IP is shown in Figure 14.

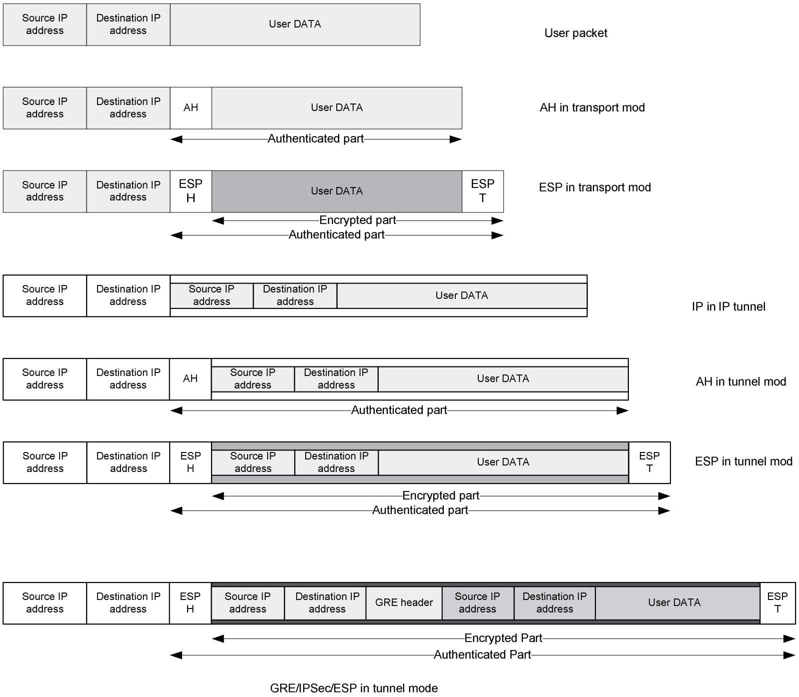


Figure 14. Encapsulating of IP packets in IPSec protocol

In transport mode AH inserts appropriate packet header between the IP header and its data part. It should be noted that in the process of authentication all elements of the packet are included: data content of the packet and appropriate fields of the IP header. In tunnel mode of operation the process of authentication is on the entire IP packet, along with the entire IP header. Moreover, in this case a new IP packet is formed where after new created IP packet header, AH header is placed and afterwards entire authenticated IP packet as the data part of the newly created packet is added.

ESP protocol in transport mode of operation performs encryption of the data part of the packet. As a result, it adds appropriate ESP header after the IP header of the packet, and at the end of the packet, after the data section, ESP tail is added. In tunnel mode, entire source IP packet is encrypted. At the beginning of the packet new IP header is added which is followed by the standard ESP header and its tail at the end. In this case the user IP packet is the data part of the newly created IP packet.

II.7.1. Setting up a network abstraction and defining of tunneling protocol

Considering all possible implementations for traffic tunneling through a particular transmission technology, the best solution for the architecture for network abstraction is a combination of IPSec between client and policy based router implemented in ESP tunnel mode and implementation of GRE tunnel over created IPSec tunnel. In this way complete protection of communication between the two ends of the tunnel in the form of data traffic encryption implemented on the IP network layer of corresponding radio transmission technology is performed. In the same time implementation of the GRE protocol over the IPSec tunnel enables complete abstraction of the network layer and is capable of transmitting non-typical IP traffic through the tunnel, as well as proper transmission of multicast and diffuse traffic which could not be transmitted if only IPSec tunneling is used. This implementation of tunneling of IP packets enables simple transfer of data traffic through IPv4 and IPv6 access networks simultaneously. Packet format of implemented tunnel using IPSec and GRE is given in Figure 15.

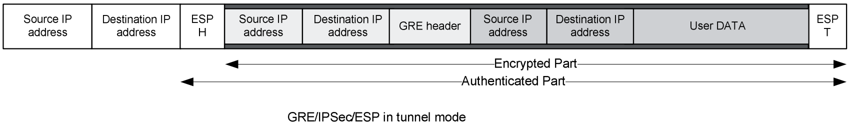


Figure 15. Format of the packet when implementing GRE over IPSec tunnel

Considering the fact that setting of IPSec protocol requires extra protection, such as public and private keys in the process of establishing of IPSec tunnel, certificates that each client receives during its registration by the policy router for the appropriate service are used. These certificates are the basis of the overall safety within the proposed architecture, and their management will be explained in more detail in the following parts related to security.

II.8. Processes of authentication in a new network architecture

One of the functionalities of the network architecture is the process of verification of the users who have the right to use the relevant services. Such verification processes is performed through user authentication and authorization processes by policy based router. DIAMETER protocol is the most sophisticated protocol for this purpose within the 3GPP and it is defined as a basis for process of authentication and authorization in this architecture. Overall security architecture besides choosing which protocol will perform authentication and authorization, also means defining the authentication parameters, so the first stage in the analysis of the security architecture embrace analysis of authentication protocol and its implementation in the new architecture. Given the fact that the first network service realized by the user, sets the basic network connectivity, whose access is regulated by basic access control mechanisms related to the characteristics of the access technology, is accomplished through data link level ("data link") from the corresponding transport radio technology as the first stage in the process of providing basic network (IP) connectivity.

These way network IP interfaces of transport technologies that represent the heterogeneity of radio access network are formed. New abstraction layer should be set over this network interfaces that will perform proper routing of user packets through appropriate technologies. In this process, the first step is the authentication and authorization of network technologies by policy based router. Given that there is already established IP connectivity over radio access networks, the processes of user authentication and authorization of each radio access technology will be realized across established IP connectivity. To satisfy this demand separate protocol for access control should be used that will provide transfer of authentication messages over an IP network. Realizing this need IETF formed a working group tasked to develop a protocol for authentication which will transmit authentication messages regardless of the transport layer. As a result of their work they created the protocol that was called PANA (Protocol for carrying Authentication for Network Access) [33]. Given that communication between client and policy server is IP based this protocol was the most suitable choice for realization of the authentication procedures between the user terminal and policy based router. Authentication chain continues to use the DIAMETER protocol for communication between PBR and customer base. The following sections describe the characteristics of the authentication protocols and mechanisms involved in the authentication process, within the defined MIM (Mobile Identity Management) module of the client and PBR that is responsible for the security procedures. Additionally, this section gives overview of additional processes and activities that are accomplished through MIM module and that are crucial for realization of the authentication and authorization processes.

II.8.1. DIAMETER protocol

DIAMETER protocol is a protocol originally designed for authentication, authorization and accounting (AAA) purposes. It represents an evolutionary step forward in terms of its predecessor designed for this purpose, RADIUS protocol. The name of the protocol is derived from the name of its predecessor, taking into consideration that the diameter is equal to twice the length of the radius. RADIUS protocol is commonly used, and is the most successful AAA protocol for providing services in fixed line (dial-up) systems, as well as wireless cellular (CDMA2000, UMTS) systems. It is a basic AAA protocol in GPRS networks as well.

DIAMETER protocol is designed to overcome the shortcomings of the RADIUS protocol. For example, DIAMETER protocol supports improved techniques for dealing with errors, ensures reliable transmission of messages, provides transmission of messages with a lot of information, introduces enhanced security, it is easily extended with other functionalities, provides flexible detection of other DIAMETER entities, etc. Moreover, and contrary to the RADIUS protocol, DIAMETER protocol defines full specification of proxy entities that are located between the two ends of DIAMETER communication. At the same time DIAMETER is designed to provide a seamless migration and compatibility with RADIUS. This can be seen from the structure of DIAMETER messages which same as the RADIUS messages contain a collection of AVP (Attribute Value Pairs). Generally, 3GPP uses the DIAMETER protocol on great number of its interfaces, however it should be noted that 3GPP in its architecture also uses the RADIUS protocol but only on Gi/SGi interface and therefore does not have significant past in using RADIUS, and at the same time does not hold any mortgages regarding its usage. Due to these reasons DIAMETER protocol is used on all other interfaces.

DIAMETER protocol is designed by defining a basic standard protocol and additional protocol extensions called applications. The basis of the protocol is described in basic DIAMETER standard RFC 3588 that specifies the minimum requirements for DIAMETER implementation and includes several DIAMETER messages called - "DIAMETER commands" as well as additional attributes "AVP-and" that are transmitted through mentioned commands. Extensions (called "DIAMETER applications") are created over basic DIAMETER protocol to support specific requirements. Applications can define new DIAMETER commands as well as new attributes "AVP's" if they are required for proper operation of applications. Description of the structure of the DIAMETER protocol is presented in Figure 16.

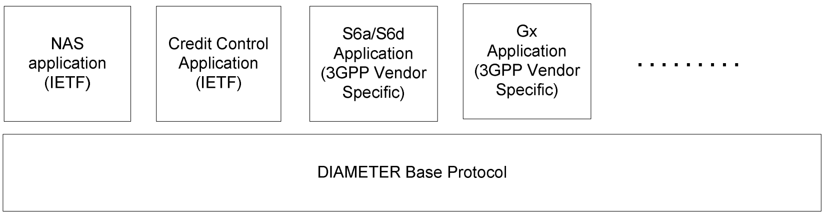

NAS application (IETF)
Credit Control Application (IETF)
S6a/S6d Application (3GPP Vendor Specific)
Gx Application (3GPP Vendor Specific)
....
DIAMETER Base Protocol

Figure 16. Structure of DIAMETER protocol

According to this structure, DIAMETER protocol application is not a program or application in the usual sense, but rather it is a protocol based on the DIAMETER. In this way applications (DIAMETER-based protocols) are taking advantage of the general features of the basic DIAMETER protocol. Also, it is important to note that applications can be based on existing, pre-defined applications. In this case, the new application inherits defined DIAMETER commands and corresponding AVP’s from base application and it can use the new identifiers and can add new APV’s that may or may not perform modifications to the procedural conditions and executions of the protocol in accordance with its needs.

Flexibility of DIAMETER protocol can be seen in Figure 16, presenting several standardized IETF DIAMETER applications used in different network architectures, and additional non-standard applications specific for each manufacturer. Manufacturer in this context does not refer only to the one that develops DIAMETER applications, but to anyone (eg. organization or company) that has received identifier for implementation of DIAMETER application by IANA. This example already exists in 3GPP, where in already defined applications for some interfaces (S6a, S6b, SWa, SWx etc.) certain new applications specific to certain manufacturers are defined. In many cases, 3GPP specific applications of certain manufacturers are based on existing DIAMETER applications, already defined by the IETF. This flexibility and expandability of the protocol makes it commonly used protocol in new network architectures.

Network entities that have implemented the DIAMETER protocol play a specific role in the network architecture. DIAMETER protocol defines three types of network entities, depending on the role of the client, server and agent. The role of the given network entity depends on its implementation within the network architecture. DIAMETER client represents a network entity that requires a service from a given DIAMETER server that starts the communication or DIAMETER session with the server. DIAMETER server is a network entity that provides certain DIAMETER service of specific DIAMETER application. DIAMETER agent is a network entity that provides a degree of flexibility in network architecture. They can be used in scenarios where different parts of the network architecture are administered by different administrative groups, such as roaming scenario. Despite this process they can play certain role in the process of routing of DIAMETER messages or in process of aggregation of DIAMETER messages distributed to a particular domain. Agents receive messages, examine the final destinations and perform their routing towards destination. This functionality can provide proper load balancing to DIAMETER servers and can simplify network configurations. Certain agents can perform additional processing on messages with specific application purposes. There are generally four types of agents: relay agent, proxy agent, redirect agent and translation agent. Unlike RADIUS where there is only one type of agent, proxy agent.

Relay agent is used to redirect DIAMETER messages to a particular destination based on the information contained within the message. The agent should have implemented basic DIAMETER protocol but does not have to understand appropriate DIAMETER application.

Proxy agent is similar in its function to relay agent, with the difference that it can perform additional processing of DIAMETER messages, for example, implementing certain policy rules. Given the fact that proxy agent can perform modification on DIAMETER messages it should understand DIAMETER service and DIAMETER application whose messages it needs to change or proxy.

Redirect agent provides a routing function, for example to ensure the resolution of domain name in a given server. Redirection agent do not forward received messages to the destination but to received messages replies with new DIAMETER message to the entity that sent the message. This message contains information that helps the initiating entity to send the message to the right destination, but this time directly to the server. Thus, the agent for the redirection is not in the path of the new DIAMETER message.

Translation agent performs translation between the DIAMETER protocol and other protocols. Typical example of such an agent is translation agent between RADIUS and DIAMETER protocol, which can be used in the migration scenario. Given that TCP and SCTP are connection-oriented protocols, before you send any DIAMETER command it is necessary to establish a connection between both ends. Both protocols provide reliable traffic. This is significantly better compared to the RADIUS protocol which uses UDP as a transport protocol and traffic between network entities is connectionless and is unreliable.

DIAMETER connection between the two entities should be different from the DIAMETER session established between DIAMETER client and DIAMETER server. While DIAMETER connection represents the establishment of transport layer between two network entities, DIAMETER session represents a logical connection that describes applicative associations between client and server (even including certain DIAMETER agents) and it is being identified through session identifier. Description of DIAMETER connection and session is given below in Figure 17.

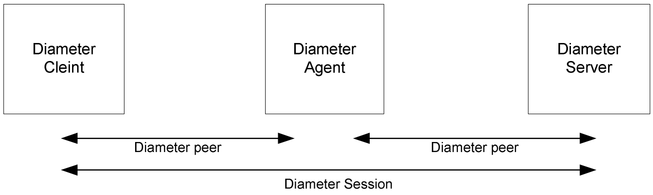


Figure 17. DIAMETER session and connection

DIAMETER messages are protected using TLS (Transport Layer Security) or IPSec. The basic specification of the DIAMETER protocol requires mandatory support of IPSec while support for TLS is optional. Security protection is provided on connection level between all network entities involved in the DIAMETER communication in step levels.

In 3GPP environment NDS/IP is used as a general framework for all-IP based control signaling, including the DIAMETER protocol. Therefore there is no need of providing additional security framework between DIAMETER network entities in such an environment.

In order to maintain the flexibility of the DIAMETER protocol in terms of application flexibility, security and in terms of dynamical change of features, two network entities that realize DIAMETER connection exchange their capabilities. This exchange allows each entity to learn the identity and characteristics of the network entity with whom it establishes connectivity (version protocol types of supported applications, supported specific attributes of a given manufacturer, supported security mechanisms, etc.).

As mentioned before, the network entities designated as DIAMETER agents can participate in the routing process of DIAMETER commands towards the final destination, DIAMETER server. In the routing process of the commands, DIAMETER agent typically performs routing of messages appropriate to their destination domain, and the application that is being used. This routing can specify different destination in accordance with the application identifier. Each DIAMETER network entity maintains a list of supported domains and known DIAMETER entities as well as their capabilities (ex. supported applications). The agent can resolve domain names in specific DIAMETER server addresses. It can be used to aggregate requests from different sources routed towards a specific destination domain. This position allows the agent to play the role of a centralized entity for routing. Each

DIAMETER entity must be able to find nearby DIAMETER entity. In the RADIUS protocol, each RADIUS client / proxy must be statically configured with information about neighboring RADIUS servers and proxy entities that needs to communicate with. This can cause great burden on network management system to maintain consistence of these configurations. DIAMETER protocol supports static configuration of neighboring DIAMETER entities, but in addition it supports dynamic configuration using DNS. In this case DIAMETER clients rely on information obtained by analysis of the domain name and DIAMETER applications and the level of security in order to determine an appropriate DIAMETER entity to which the message can be forwarded. These obtained entities and routing information are stored locally and used in decision-making during the process of routing of the messages. Basic DIAMETER protocol includes mechanisms to support transport failover between transmission entities, for example by using the messages for detection of established connection and detection of transport errors. Description of the format of DIAMETER messages is shown in Figure 18.

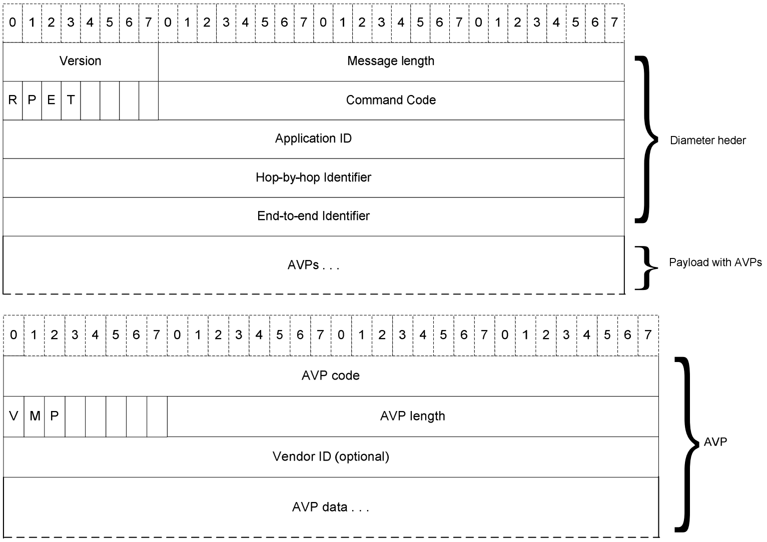


Figure 18. Format of the DIAMETER protocol messages

DIAMETER messages are called commands. The content of DIAMETER commands comprises of DIAMETER header followed by a number of AVP-s.

DIAMETER header contains a unique code for identification of the command followed by description field of its purpose. Actual application data is transmitted over corresponding AVP part. Basic DIAMETER protocol defines a set of commands and attributes needed for normal operation and in addition to that each application can define additional commands and attributes specific to its functionalities. The basic protocol, for instance, defines a set of commands and attributes that can be reused as objects in the process of defining new attributes. In addition to defining new attributes in a given application it is possible to define new DIAMETER commands which makes DIAMETER protocol very flexible and extendable and as such allows the design of various applications that can meet the requirements of 3GPP.

The "Application ID" field identifies the application for which DIAMETER message is intended for. "Hop-by-hop identifier" field is used for matching requests with responses. "End-to-end identifier" field is used for detection of duplicated messages. Each pair of attributes has its own unified AVP code. If the pair of attributes is specific for a given manufacturer, beside the basic fields, new unique field - "Vendor ID" is added to define the manufacturer.

II.8.2. EAP - Extensible Authentication Protocol

EAP represents authentication framework that can support different authentication methods. EAP protocol performs a kind of abstraction that provides an adaptation between any authentication method and any access technology without the need of their tight integration. The basic idea behind the design of the EAP protocol is associated with the ability to transfer messages from the EAP protocol through embedded technologies, that is to say, when a network has the ability to transfer EAP packets then it can use all available authentication methods implemented as EAP methods. EAP protocol is not authentication method by itself but common authentication platform which can be used for implementation of specific authentication methods. Today there are more than 50 different authentication methods which can be used over EAP protocol, and their number is expected to grow, without needing to change the transmission technologies.

The basic EAP protocol is defined in RFC 3748. It describes the format of the EAP packet and its basic functionality, such as a process of negotiation of authentication mechanism. It specifies several simple authentication methods, such as methods based on one time password or “challenge-response” authentication methods similar to CHAP. In addition to the authentication methods defined in RFC 3748, it is possible to define additional EAP methods. These EAP methods can implement other authentication mechanisms and/or use other authorization mechanisms as mechanisms based on public certificates as part of the PKI (Public Key Infrastructure) or U (SIM) cards. Several of the IETF standardized EAP methods are defined below:

- EAP-MD5 is defined in RFC 1994. This mechanism is based on CHAP (Challenge Handshake Authentication Protocol). Operation of this mechanism is based on a shared secret key that is known both by the client and the server that are in the process of authentication.
- EAP-TLS is based on TLS (Transport Layer Security) and defines the EAP method for authentication and key derivation based on public certificates. EAP-TLS is defined in RFC 5216th
- EAP-SIM is defined for authentication and key derivation using a GSM SIM card. EAP-SIM enriches basic GSM SIM procedure by adding support for mutual authentication. EAP-SIM is defined by RFC 4186.
- EAP-AKA is defined for authentication and key derivation using UMTS SIM card and is based on UMTS AKA procedure. EAP-AKA is defined in RFC 4187th

In addition to existing EAP authentication methods, non-specific EAP methods also exist within the corporate WLAN networks developed by certain manufacturers. (Ex. EAP-LEAP, etc.).

According to the architecture of the EAP protocol communication is based on three network entities:

- EAP client: represents a network entity that requires access to network resources. A typical example is the user terminal. In EAP implementations in WLAN (802.1x), it is called supplicant.
- Authenticator: an entity that performs access control, such as a WLAN AP or ePDG.
- EAP server: represents background authentication server that provides the authentication service on authenticator. In EPS architecture, example of such a network entity is a 3GPP AAA server.

EAP protocol architecture is presented in Figure 19 below.

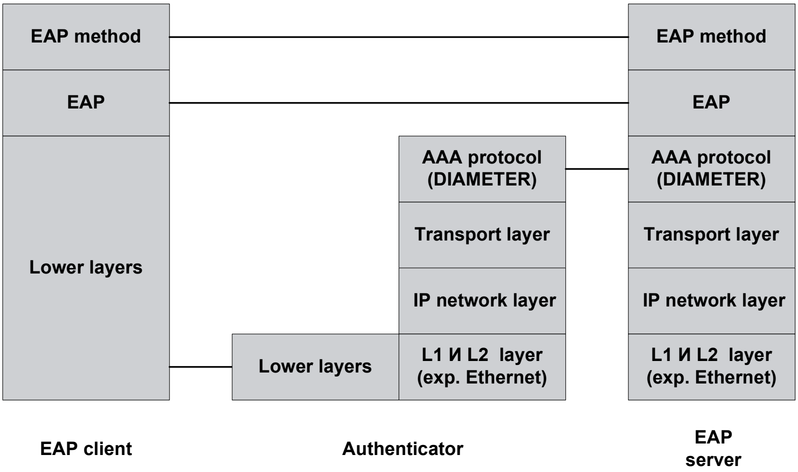


Figure 19. Architecture of EAP protocol

EAP protocol is commonly used in processes of user access control before IP connectivity is set, and is performed between the terminal equipment and network equipment. Between the user equipment (EAP Client) and authenticator, EAP messages are typically transported through the lower data layers ("data link"), such as PPP or WLAN (IEEE 802.11) in the processes that precede the setting of IP connectivity. In this process EAP messages are encapsulated directly in the protocols of the lower layers. The way that this is conducted is described in various EAP specifications. For example, RFC 3748, describes the EAP implementation through the PPP protocol, IEEE 802.1x describes the implementation of EAP over IEEE 802 links like WLAN, etc. EAP protocol can be used for authentication with IKEv2, and in this case EAP messages are transported through the IKEv2 protocol and IP. EAP messages between the authenticator and the EAP server are typically transmitted through an AAA protocol (RADIUS or DIAMETER). EAP communication between the EAP client and the EAP server is transparent to the authenticator, and therefore the authenticator does not have to support the specific EAP methods used in communication, but should transparently redirect these messages between both ends of the communication. Description of the procedures in the EAP authentication process is given in Figure 20 bellow.

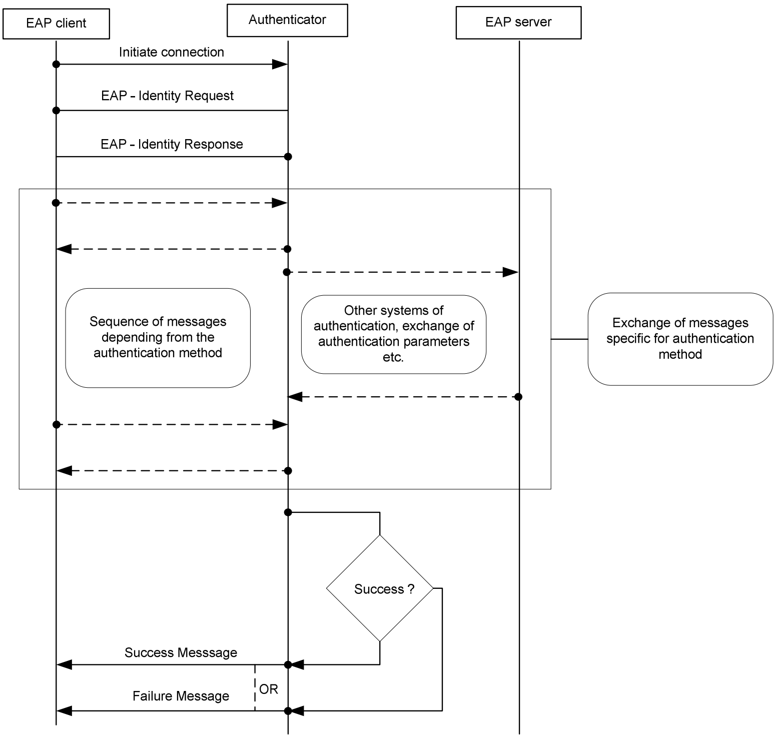

EAP client
Authenticator
EAP server
Initiate connection
EAP - Identity Request
EAP - Identity Response
Sequence of messages depending from the authentication method
Other systems of authentication, exchange of authentication parameters etc.
Exchange of messages specific for authentication method
Success ?
Success Messsage
OR
Failure Message

Figure 20. Messaging under EAP authentication framework

Authentication with EAP typically begins with a process of negotiation of EAP method to be used in the process. Once EAP authentication method is negotiated and accepted between network entities, a process of exchange of EAP messages between the EAP client and the EAP server is initiated, through which the authentication is started. When the process of authentication is completed, EAP server sends the EAP message to the EAP client, which confirms or denies the success of the authentication process. Authenticator is informed about the result of the authentication process through the AAA protocol. Based on this information authenticator can provide user access to the requested access network, or can continue to block access. Depending on the chosen EAP method, EAP authentication can be used in the derivation of keys by the EAP client and the EAP server. Such keys can be transported through the AAA protocol from the EAP server to the authenticator. After completing this procedure and getting derived EAP keys on the side of the terminal (EAP Client) and authenticator, the network elements through which transmission of data is performed, the process of derivation of transport keys to protect access link transmission data can begin. In the scientific community there are a number of different research and development activities related to the EAP protocol. One are set in the direction of development and design of a new EAP authentication methods that aim to support new authentication schemes developed over the development of wireless technologies, while others are associated with the development of other new transmission layers for transmitting of EAP messages. In line with development of new method of authentication, we can highlight the example of EAP-SAML, while in the field of development of new transmission techniques for the transmission of EAP protocol, the development of PANA (Protocol for Carrying Authentication for Network Access) which is a protocol for transmission of EAP protocol over any transport networks is particularly important.

II.8.3. PANA - Protocol for carrying Authentication for Network Access

PANA is a protocol designed by the IETF as a protocol that should provide transparent network authentication over data link network layer. Its goal is to ensure the establishment of any authentication protocol, over any transport technology (data link level). This goal is achieved by setting the EAP over IP as the transmission level. Along with this basic principle, this protocol provides a number of additional and powerful functionalities, such as: separate NAP and ISP authentication possibility of reuse of local security associations, fast reauthentication, and secure exchange of EAP messages protocol extensibility by introducing additional protocol messages and so on. Such placement of PANA protocol makes it suitable for use in procedures for authentication in heterogeneous networks. PANA and IEEE 802.1x are similar to each other because both protocols transmit EAP messages between the client and the network. The most important difference between them is that PANA can be used on any data link layer, while IEEE 802.1x can be used only through the IEEE 802.1 networks, and, in addition to that, IEEE 802.1x lacks additional functionalities. For standardization of PANA protocol IETF formed a working group that takes care for its development and unification. PANA protocol is set on the last IP link between PANA client (PaC) and PANA authentication agent (PAA)

PAA client is set at the controller for network access side. This access controller attempts to bridge the AAA sessions between the client and the AAA server using PANA on one side (the client) and DIAMETER / RADIUS on the other side (to the AAA server). PANA platform, despite these network entities also defines other network entity called EP (enforcement point). EP control access in a manner that prohibits access of unauthenticated users to the network resources. This process is enabled by packet inspection and filtering of network packets. Filtering can be based on simple parameters such as source and destination address, but these methods are not adequate for use in multi access wireless networks, and that's why IPSec-based control is a typical implementation in these systems. EP entity should be placed at key locations in the network architecture in order to allow full control of the traffic in both directions. Example of this EP can be AP in WLANs. Otherwise, when there is no separate entity between the client and the access router, access control can be implemented by the access router.

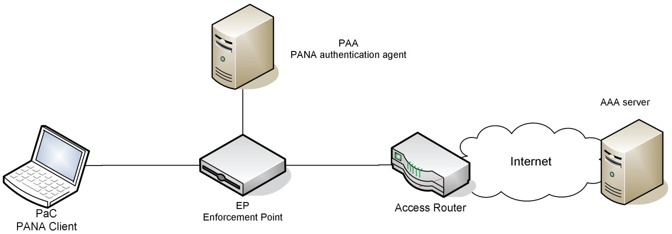


Figure 21. Overview of the architecture of PANA protocol

In most networks, the EP and the authentication agent (PAA) are collocated within a single network entity, but despite this PANA protocol allows separation of these two functions in the architecture in different entities as shown in Figure 21. This separation allows control of more EP by a single PAA client and needs to establish a protocol that works between EP and PAA network entities. For this purpose PANA - IETF working group has proposed SNMP protocol to support the required communication between EP and PAA entities. This functionality of PANA protocol allows great flexibility in the organization of the transmission network, depending on the specific requirements and needs of transport network. It has to be mentioned that PANA represents the lower transmission layer of EAP protocol as seen in Figure 22.

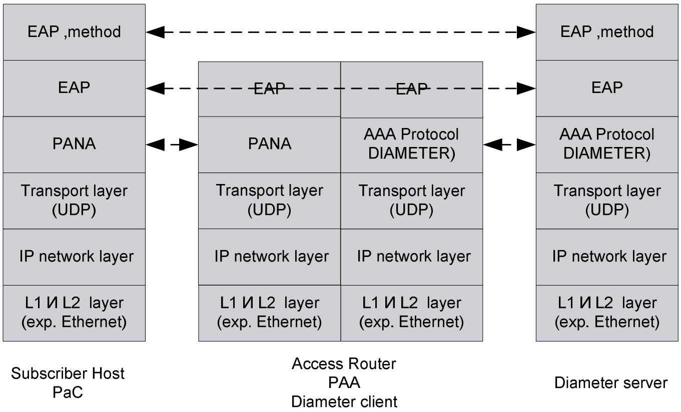


Figure 22. Protocol communication diagram by PANA protocol

PANA protocol in IP networks is defined as UDP based protocol that operates between two IP-based network entities on the same IP link on UDP port 716. It provides transmission of messages in a specified order as required by EAP specification. Transport message flow is shown in Figure 23. Initiation phase of the protocol consists of exchange of a series of messages with requests and responses. Some of the transmitted messages carry information between the client and the network, while others are used to manage the entire PANA authentication session.

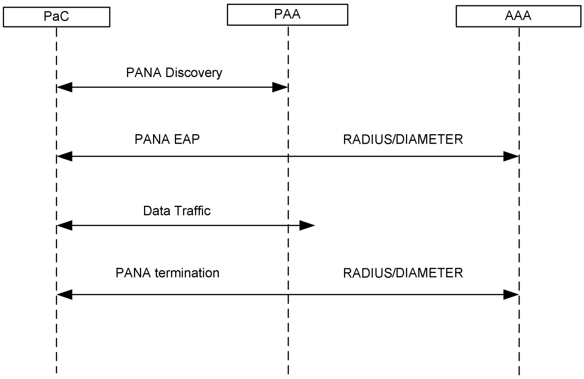


Figure 23. Message flow of PANA protocol

Discovery phase involves two possible scenarios for discovery of network entities. The first concerns the discovery initiated by PaC client that sends a message to detect PAA on access links, while the second refers to the discovery of PAA which begins the process of discovery after the process of connecting of PaC on access link. This process of detection of network entities can be omitted and process can move directly to determine their identity in case when PaC has information about the IP address of the PAA entity. After this process is finished authentication process goes on with exchange of start PANA message, which also marks the beginning of the PANA session. After this, the process continues with exchange of series of more PANA authentication messages. These messages simply transport EAP data between the EAP client and authenticator over PANA protocol. Given that PANA is a transmission protocol structure to carry EAP protocol, it does not perform further analysis of the contents of EAP messages, only messages from EAP protocol that have importance for PANA messages, indicating successful or unsuccessful completion of the EAP (EAP-success and EAP-failure). These messages mark the end of the authentication phase after which further EAP messages have to be transmitted within the PANA-bind message. If authentication is successful, the message simultaneously imposes a common understanding of the identifier of the device (if the device is identified by its MAC or IP address) and associated level of protection of exchanged packets. Agreed identifier of PaC will be handed over to the EP for performing access control in the next phase. Meanwhile the two entities can decide whether to use the basic link-level encryption or IPSec for additional cryptographic data protection level of exchanged packets. This type of mechanism can be enabled only if the EAP methods used in the process of authentication provide derivation of cryptographic keys. Received keys are used to generate PANA SA (PANA security associations) that are used to protect the exchange of multiple consecutive PANA messages. Once PaC client is authorized it can start with normal IP communication via the EP to the protected network. During this phase PaC and PAA can perform continuous checking of the connection state, realized by asynchronous sending of "PANA-Ping" messages. These messages are good for the detection of broken connection between the two ends. PANA protocol defines duration of authenticated sessions. After this time, if PaC wants to continue the connection between the two entities they need to enter a new cycle of authentication (reauthentication). If for some reason the customer leaves the network or PAA wants to end the connection over PANA protocol, "PANA-terminate" message is being sent that indicates the end of the established PANA session. PANA protocol is executed in the same order regardless of the environment in which it operates. In less secure environments it is expected to choose EAP methods that provide mutual authentication and generate cryptographic keys in order to protect the communication. Moreover, these keys should be bind to data traffic which is realized by an additional protocol for their exchange that will follow the successful PANA authentication. [34]

PANA protocol allows IPSec-based access control in a way that helps the IPSec protocol in the creation of IPSec security associations. PANA protocol generates cryptographic keys upon completion of the EAP protocol for creating PANA security associations, but they cannot be used directly to create IPSec security associations. This relationship can be used as a basis for "pre-shared secret" for generating dynamic IPSec associations. This approach leads to the use of IKE protocol for reuse of PANA security associations for IPSec associations for IPSec-based access control. The keys are obtained by derivation of the PANA security association and as such are delivered to IKE protocol. This new IPSec associations are used further to create a tunnel between the PaC and PAA, or EP entity for providing of authenticated or encrypted data transport.

II.8.4. Concept of authentication in a new network architecture with use of predefined authentication mechanisms

Process of authentication and authorization based on PANA protocol as forefront for formation of authenticated IPSec tunnel is the basis of the design of the proposed architecture for interworking among heterogeneous networks in this book. As mentioned in the main text, the first step in the process of achieving IP connectivity through access technologies is their authentication followed by the formation of their abstraction by creating of IPSec\GRE tunnel. For authentication protocol in this case it is best to use EAP and as authentication method on top of it is best to use TLS. As a transfer protocol for authentication between client and policy-based router in this case PANA protocol will be used where user terminal will be treated as PaC and policy-based router as PAA. In the background, policy-based router should achieve connectivity with CPH module (server) using DIAMETER as authentication protocol. Therefore, in the authentication process the user terminal acts as EAP client as long as CPH module (server) acts as the EAP server. EAP session is established directly between these two entities. After successful authentication and authorization of user using the PANA protocol as described above and by using of IKE protocol, the process continues in the stage of formation of the IPSec tunnel, and creation of the network abstraction through a GRE tunnel between the two ends of the communication, the user terminal and PBR. Use of EAP-TLS as authentication protocol is due to the concept of security set in the new architecture. The security module (MIM) by the client and PBR poses public certificates that consist of public and private cryptographic keys. These certificates are issued by the organization that controls the service it offers to customers. There are various ways how this can be done, but mainly it is based on setting up a certificate-based security infrastructure (PKI). There are many ways a customer can get this certificate. Considering the dynamic of modern life, the best way for the operators and customers to make this is by using the Internet itself. Namely, if the client is willing to use the service it may require to purchase it, this can be done by approaching the web site of the operator who offers the service and after it, will make the electronic payment on the same web site in a secure manner (over SSL) the client will be offered the option to download the certificate to its user terminal.

With the help of such infrastructure each customer is associated with a certificate that has been issued for its use in this architecture. Such certificate (X.509) has its own characteristics and consists of a number of parameters, the public key, private key, and part of the additional elements that describe its purpose, the organization which issued, and its validity (time of its generation and termination of its validity). In the process of user authentication via EAP-TLS protocol check of these certificate parameters is performed to the user who in CPH module is connected to the right customer. If its authenticity and validity are determined and also the client that is represented by this certificate is entitled to use the appropriate service it will enable establishment of a link between client and PBR for a specific technology. In this process CPH performs authentication (authorization) of access technology in order to determine whether it can be used in the interworking operation. This completes the client's authentication and authorization phase. Details are shown in Figure 24.

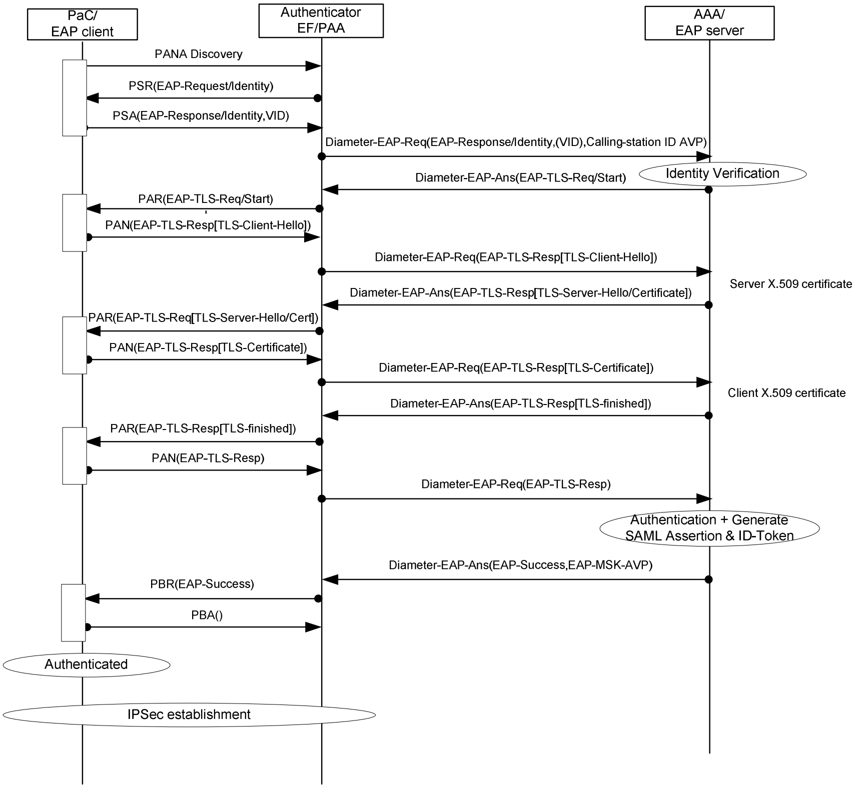

PaC/ EAP client
Authenticator EF/PAA
AAA/ EAP server
PANA Discovery
PSR(EAP-Request/Identity)
PSA(EAP-Response/Identity,VID)
Diameter-EAP-Req(EAP-Response/Identity,(VID),Calling-station ID AVP)
Diameter-EAP-Ans(EAP-TLS-Req/Start)
Identity Verification
PAR(EAP-TLS-Req/Start)
PAN(EAP-TLS-Resp[TLS-Client-Hello])
Diameter-EAP-Req(EAP-TLS-Resp[TLS-Client-Hello])
Server X.509 certificate
Diameter-EAP-Ans(EAP-TLS-Resp[TLS-Server-Hello/Certificate])
PAR(EAP-TLS-Req[TLS-Server-Hello/Cert])
PAN(EAP-TLS-Resp[TLS-Certificate])
Diameter-EAP-Req(EAP-TLS-Resp[TLS-Certificate])
Client X.509 certificate
Diameter-EAP-Ans(EAP-TLS-Resp[TLS-finished])
PAR(EAP-TLS-Resp[TLS-finished])
PAN(EAP-TLS-Resp)
Diameter-EAP-Req(EAP-TLS-Resp)
Authentication + Generate SAML Assertion & ID-Token
Diameter-EAP-Ans(EAP-Success,EAP-MSK-AVP)
PBR(EAP-Success)
PBA()
Authenticated
IPSec establishment

Figure 24. Exchange message flow in EAP-TLS over PANA and DIAMETER

There are additional techniques that enrich the processes of authorization and authentication. One of these protocols is EAP-SAML. SAML (Security Assertion Markup Language), which is an XML-based standard for exchanging of authentication and authorization data between security entities, that is between the entity that poses the identity and the one who provides the service. This standard is a product of the OASIS (Organization for the Advancement of Structured Information Standards). The use of this EAP method provides an opportunity for exchange of additional policies between the authentication server (CHP module) and the authentication client (MIM module of the terminal). It can be used for transfer of basic policies as well as for transfer of the initial McIP address of the client abstract layer.

Concept for providing application QoS/QoE as a basis for policy routing

Given the fact that the new concept of next generation mobile networks implies heterogeneity in the nature, and the fact that the underlying network protocol over which client services are realized is IP, there is an essential need for providing applicative QoS that should meet the QoS requirements of embedded IP networks. In order to classify the applicative QoS requirements it is necessary to determine the basic QoS parameters of IP transmission network having in mind the way that a QoS parameters for given application can be classified for transport over transmission technologies. These parameters are defined as main and basic performance indicators of the quality of the transmission technology in terms of IP transmission network. In the following part of the text special attention will be placed on the characteristics of the network, their qualitative parameters and the method of measurement, and their contribution in providing the ultimate applicative quality of service.

II.9. Introduction to basic concept of related-performance indicators

Key performance indicators or (KPI) represent measures that are of particular importance for the customer service performance. As a first step in the analysis of quality of service of customer services is the definition of basic indicators, and the second step is determining how to measure them [35], [36], [37], [38], [39], [40]. In the telecommunications world, there are three major groups of KPI:

- Basic indicators for network availability.
- Basic indicators for network sustainability.
- Basic indicators for quality

Availability is related to the user ability to set a customized service and access the radio resources. Sustainability provides information about how good the quality of service is (e.g., sustainable data rate), while quality can be determined by the success rate of the particular service.

II.9.1. Key-performance indicators

In packet based networks, the links between network performance and customer service performance in most cases are not seen as easily as in transmission technology based on circuits switching. This difficulty is mainly due to multilevel structure that characterizes packet communications. Performance problems from lower layers, such as unreliable link, can be reflected on upper layers as increased delay. This performance of the links among various applications, causes various degradations depending on the change of different parameters of packet layers of the links. An example of this can be given by an overview of two applications, Web and MMS. Web applications suffer performance degradation when the delay is very large, while the MMS application can endure delay of ten seconds without significant degradation of the performance of the user service and their satisfaction. Generally there are a number of indicators that affect the performance of the service from end to end and thus directly affect the quality of the end-user experience. These performance indicators in view of their importance are the basis of all measurement systems, through which you can perform appropriate qualitative assessment of transport technologies in terms of application performance. [41], [42], [43], [44]

This group of performance indicators includes: packet delay (RTT – delay of packets in both directions, delay in access establishment and jitter), a flow rate of packets and packet speed, reliability (number of lost packets, bit error rate and packet error rate) and availability.

II.9.1.1. Packet delay

In telecommunications, especially in data transmission technologies we can distinguish three different types of delays that are commonly considered: packet delay in both directions (Round Trip Time - RTT), access delay and Jitter. For data links having no time of arrival constraints of packets related to the service, the first two delays have a significant role in determining the performance of services, while for real-time services, such as speech and video, jitter is of particular importance.

Packet delay in two directions (Round Trip Time - RTT) represents the time between sending the packet by the network entity and the time until it is received back. RTT time greatly depends on the distance between the two entities between which it is measured (geographical distance), and the delay that occurs at each step in the transition of the packet from one to another network which is located on the transmission path. In RTT measurement there is a limit which is affiliated with asynchronous and asymmetric links, where one-way delay of packet may be different compared to the same in the other direction. This case cannot be detected by classical RTT measurement because its measurement is limited to measuring the time between sending an echo request until its receipt by the same network entity. This is presented in Figure 25, where it can be seen that the RTT time represents the sum of the times of propagation of the packet in one direction T1 and response time required to turn back in the other direction T2. Delay as a result of processing of the echo request in client or server network entities is not taken into account in this picture.

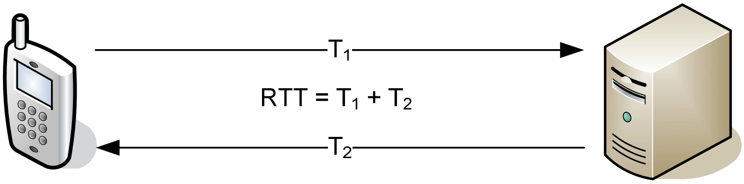


Figure 25. Packet delay in two directions – RTT

The simplest and most common way of measuring of packet delays in both directions is by using "Ping" application, which sends ICMP (Internet Control Message Protocol) - "echo request" messages/packets to destination entity and listens to the answer in form of an "echo response" packets. Some form of Ping application exists in all major operating systems (Microsoft Windows, Linux with its distributions and UNIX with its variants). In addition to the calculation of the time delay in both directions, Ping application estimates the packet loss in the process of RTT measuring as well. However, the preferred case where application uses ping is to check the connectivity of a network entity to a given destination in a local area network or Internet. When measuring the quality parameters for certain access technologies it can be generally said that the smaller the RTT time, the better connectivity of certain access technology. However, different services have different requirements for the size of the RTT in order to meet user expectations and requirements for quality of service. Generally we can say that the RTT size of 30 ms, measured at the access technologies can be considered as a good time, while the RTT size of 3 sec. will bring great degradation in service performance. Transmission protocol such as TCP, which relies on a mechanism for acknowledgment of received packets before sending of the next packets, is strongly affected by the latency of end to end transmission technology. Packet delay is mainly due to the delay in propagation through the network. Simply put, it represents the time required to transmit the signal from one place to another through the transmission technology (e.g. wire). Transport delay is a result of the transmission medium, e.g. longer packets will be transmitted longer than shorter packets. Delay is also generated in the routers that inspect packet headers and perform change of the value in the TTL field.

Delay in the access establishment, represents a delay that occurs in the process of communication establishment between the mobile client and the core network. When switching on the mobile phone and opening the browser for the first time, the user feels a certain delay that occurs in the process of establishing a GPRS connection or process while mobile client realizes basic network connectivity to the core network through transmission technology. This delay is normally in the range of a few seconds and as such should be kept in that range, as access delay greater than this (in range of minutes) is unacceptable and makes the service unusable.

Jitter represents the variance in packet delay in the transmission of data in a given packet network. All IP transmission technologies have a jitter and it plays an important role in real-time applications and services because of the need for arrangement of packets in proper order so that information is received in the same order as it is generated.

II.9.1.2. Packet flow speed and packet speed

The speed with which the network is able to send and receive data or data packets is called data speed, or bandwidth speed. Throughput in other hand represents the bit rate and it is limited by the capacity of the network channel. It is most commonly measured as the number of bits per second transmitted through the transmission segment by defining 1 kilobit as 1000 bits, as opposed to the size of the data file that is commonly measured in bytes and where 1 kilobyte represents a value of 1024 bytes. Potential or theoretical speed of particular packet network is called a data flow while packet speed represents a specific number. The reason that the data flow is more frequently referenced than the data speed is because it is easier to calculate. The data speed is more difficult to calculate because it depends on a number of different variables, such as packet loss and the transport protocol. This means that the user data rate is higher using UDP over IP rather than using TCP over IP, which is the reason that services which require higher speeds and real-time operation or smaller delays are using UDP as a transport protocol.

There is often a difference between what can be measured as average packet speed, a speed that can be expected in the process of transferring the data file and the maximum peak bit rate that can be achieved in certain short time intervals (in particular data bursts).

II.9.1.3. Confidentiality

The degree of error represents the probability that packets are lost or received with a certain error. Confidentiality on data and network layer transmission directly affects data speed. Packet losses in the network layer when the TCP is used as the transmission technology will cause retransmission and therefore will cause a reduction of the data throughput of the link. Packet loss in video streaming and UDP transport protocol will cause a reduction of image quality and occurrence of interruptions in the transmission of the image.

In IP-based communications two most used transport protocols are TCP and UDP. TCP provides a reliable transport and achieves reliable communication between two network entities. If one of the segments is lost in transmission or it is removed due to an error control mechanism within the TCP, then TCP transport protocol will perform retransmission of the lost/erroneous segment. Other mechanisms within the TPC as a flow control mechanism and congestion control mechanism will reduce transmission speed of TCP stream. Unlike TCP, UDP transport protocol continues to send data stream with the same data speed regardless of network conditions and the fact whether the same information is delivered to the other end of the link.

II.9.1.4. Availability

The probability that a particular service is accessible and available to the requirements of the end user is called service availability or simply availability. Availability is usually affected by the availability of the network and the stability of the system and applications that are the subject of analysis, which is part of the infrastructure of the service provider.

II.9.2. Classification of key-performance indicators

According to type and way of measurement of main-performance indicators they can be divided into two categories:

- Passive performance indicators - which are measured directly in the management systems of the networks without additional active participation. They are usually measured at intervals.
- Active performance indicators - which are continually actively measured with various tests and monitoring tools. They provide a high level of detail, usually without statistical information, resulting in frequent repeatability of measurements.

According to the focus area in terms of performed measurements, the key-performance indicators are further divided to:

- Network performance indicators. Most of these indicators are passive, because they give an image of the performance of the communication system in terms of the system itself. This includes performance monitoring processes such as radio resource management, mobility and so on. These performance indicators are usually monitored to detect certain capacitive network problems.
- Service-based performance indicators. These indicators depict the performance of certain services from the perspective of service users. They address the performance of the service and the user experience of the service, something that network indicators themselves do not provide. For example, service-based indicators for HTTP/Web search are defined in a way that they present real image of the monitored service from

customer perspective. Such performance indicators are mainly active or have appropriate combination of active and passive monitored indicators.

Most important indicators in terms of observed end-to-end service are service indicators. Network indicators portray the state of the network, and to some extent they can be linked to certain service indicators or at least to determine their subtle interdependence. Network indicators give preview of a part of the overall system that is involved in the realization of customer service and are mainly focused on one segment of the system, the network segment. This segment itself may consist of several different segments and for each of them there is a different way of measuring these indicators. For this reason, measurement of network indicators from end to end is quite difficult. If we want to make such measurement we need to implement active measurement with specific monitoring tools, similar to the way service indicators are measured.

II.9.3. What influences the service based KPI?

Factors, or indicators, that specifically affect the performance of specific service, are service based KPIs. This means that there are different KPIs for HTTP web browsing or MMS or FTP file transfer. In the following section we will address some of the KPIs for major services in packet switched networks. In fact, KPIs are defined in Chapter 5 of this book, while here we will discuss the motives that lead to KPI definitions.

FTP

For the end user to be satisfied with the file transfer protocol, some criteria may have to be fulfilled. First of all, he must be able to connect to the server within a reasonable time. When he has connected and set up the control connection, he must be able to establish a data connection. When both control and data connection is established, the reliability of the data connection is important. This leads us to the throughput. All delays and connection failures will degrade the throughput. The actual bit rate the link is able to provide while download is in progress is affected by the throughput of the link.

HTTP / Web browsing

Some of the KPIs that apply to FTP also apply here. In fact, both these services are about transmitting files. However, from end-users perspective, there are certain things that in most cases will give a better experience. For instance, some data takes longer time to load than other. If you were to wait until all data (text + pictures+ applets) were loaded before rendering the page, this will in most cases take some time, especially on low bit rate connections. If you get the text up and readable early, this will in most cases give a better user experience, than if you have to wait for a longer time, and only see the page when it is fully loaded and rendered.

Multimedia Messaging Service (MMS)

The MMS is a service similar to SMS, but with a lot of enhancements when it comes to features and content. With MMS one is able to send pictures, video clips and sound. It is transferred packet switched, unlike SMS which mostly is transferred via CS signaling. The most important thing for the user is that it is delivered. In most cases some delivery delay can be tolerated, although it is always a positive thing to get it delivered as fast as possible.

Ping

Ping has only one KPI, the round trip time, or shortly RTT. It is the time it takes for a packet of different sizes to go from one host to another and back again. It measures the latency of the network. It sends “echo request” ICMP packets to the host, and waits for “echo response” packages.

II.9.4. Factors affecting the performance of customer service from end to end

Final performances of the customer service are affected by every protocol level and network elements on the path between end points where service is set. This implies that every regular level bottom-up degrades the performance of the link. Packet flow on physical level is used as the initial value and then the proper estimation of the degradation that occurs in the upper levels is performed. Other factors that impact performance degradation on the link level flow can be divided into two groups: the degrading effects of the data link level and degrading effects of the higher levels. Degrading effects of the data level depends largely on embedded technologies through which the transmission of the information from higher levels (IP packets) is performed. In these effects are all elements that are the result of the corresponding radio resource management techniques of individual technologies, and corresponding radio protocol functions related to data transmission in wireless transmission domain.

II.9.4.1. Data Link Effects in GPRS/EDGE network

Those factors that degrade the performance depending on radio coverage, interference and resource sharing are called data link effects. The performance after these degradations is called data link throughput, and is the final throughput offered by the Radio Access Network to the upper layers. Data link throughput and latency can be calculated based only on the network itself.

GPRS is affected by the interference levels in the frequency planning and delays occurring with transmission times between BTS and BSC, Radio Resource Management (RRM) functions and radio protocol functions. From the perspective of data link effects, one can define the following performance indicators:

Coding Scheme	1 slot	2 slots	3 slots	4 slots	5 slots	6 slots	7 slots	8 slots
CS-1	9,05	18,2	27,15	36,2	45,25	54,3	63,35	72,4
CS-2	13,4	26,8	40,2	53,6	67	80,4	93,8	107,2
CS-3	15,6	31,2	46,8	62,4	78	93,6	109,2	124,8
CS-4	21,4	42,8	64,2	85,6	107	128,4	149,8	171,2

Table 1. GPRS Throughput at LLC layer

Peak throughput

The throughput delivered to the LLC layer without RLC/MAC headers depends on the used modulation and coding scheme (CS/MCS). The peak throughput given for each coding scheme in GPRS can be seen in Table 1.

Timeslot capacity

The timeslot capacity is the available throughput in a timeslot (TS) after including the effects of interference and RLC retransmissions if RLC acknowledged mode is used. There are several factors that affect TS capacity, such as radio link quality, network planning (frequency reuse) and configuration, the layer where GPRS is allocated (BCCH hopping, non-hopping).

Reduction Factor

Timeslots are shared between several connections. The reduction factor (RF) includes the fact that it is a shared medium. Network load and dimensioning conditions affect the RF and it depends on several factors:

GPRS allocation size: How many TS are reserved for GPRS and how many are shared between voice and data is important to know to prevent high GPRS blocking, and thus high RF.

CS (Circuit-Switched) load and pre-emption criteria: The priority given to CS and PS traffic is important in preventing RF. If CS traffic is given higher priority than PS, high CS load will degrade PS traffic too.

Terminal capability, Multislot class: Terminals which support several TSs are capable of getting higher bit rates from the system. A high-end phone, like Nokia N70, is typically capable of using 4 timeslots for downlink and 2 for uplink, i.e. switching between 4+1 and 3+2 as maximum number of timeslots that can be used simultaneously.

RRM scheme: The job of the RRM is to take care of minimizing the TS sharing when doing channel allocations. It ensures that the terminal is connected to the best cell.

RLC signaling

Whenever data needs to be sent through the radio interface, a Temporary Block Flow (TBF) has to be established. The TBF may cause some delay when it is being established, typically in the area 300 – 600 ms, and thus TBFs being released and established continuously may cause performance degrading. The throughput will also be affected given the fact that the RLC control blocks used for signaling shares the same radio resources as that of the data RLC blocks.

One wireless event that may affect upper layer behavior is the one originating from the mobility issue. Cell reselection will cause some delay, in the level of seconds. Various new enhancements have been introduced to lower this delay, three such enhancements are Network Controlled Cell Reselection (NCCR), Network Assisted Cell Change (NACC) and the use of Packet Common Control Channels (PCCCH) for signaling. Using these together may minimize the cell change delay too around 500 ms. RLC retransmissions will cause higher delay and jitter; this can affect upper layer protocols like TCP.

II.9.4.2. Performance of the TCP protocol in wireless networks

By analyzing the performance of the TCP protocol in wireless networks it can be concluded that there are certain differences in comparison with fixed (wired networks). Greatest impacts on the performance of these networks are by the following factors: packet delay, packet loss, variable change of the data flow, asymmetric geometry of traffic and so on.

Packet delay has a significant impact in the performance of the network that transmits TCP packets. High latency (delay) measured as large RTT time, causes slow start TCP window-slide mechanism. Exactly this is the case in most of the wireless transmission technologies, which result in a slow increase in the transmission speed in applications that run over TCP. The result of large delays is increased setup time of the TCP connection and slow exit of TCP slow start mechanism. Despite the fact that TCP refreshes the timeout value for retransmission (RTO) based on acknowledges it receives by TCP mechanism itself and based on RTT time, sudden delays due to change in radio conditions can cause problems. When this change occurs TCP suffers peak packet delay. Such delays can occur as a result of various events in the wireless domain, such as retransmission on radio control level, due to the poor quality of the link as a result of bad coverage or due to handover between cells in case of congested cell with higher priority traffic and so on. Large packet delays can cause a timeout in TCP connections, in which case the TCP mechanism assumes that the packet is lost and starts the retransmission process by slowing the transmission flow and re-initiating TCP slow start mechanism.

Second important segment which has a significant impact in performance of the TCP is the packet loss across the transmission network. In wireless networks as opposed to the wired, where packet losses are mainly due to the overloading of certain buffers in the transmission paths of the packets, packet loss are due to errors that appear in the radio transmission link. The emergence of radio link control (RLC) and confirmation of transmitted packets, provides reliable radio link, preserving the rule of delivering packets in time order and it is always used. Packets may be lost due to certain phenomena in process of handover. TCP mechanism on such losses reacts as some congestion happened in the network and therefore activates the slow start mechanism, which reduces the packet transmission rate in half.

Third segment which affects the performance of TCP is a variable data rate. Sudden change of the bit rate directly affects the TCP control mechanism. Change of bit rate in radio transmission technologies can occur for various reasons, for example, the number of users connected to a cell affects the available data bandwidth to users, the distance between the mobile client and a base station can affect the client flow due to different radio coverage, etc. Although the TCP mechanism can adapt to the speed of the flow of radio transmission technology, fast and sudden changes may lead to reduced capacity utilization of the link or lead to use of a wide TCP window. Sharp reduction of the capacity of the link can lead to packet loss and force the TCP to trigger slow start mechanism. Besides these three main reasons there are several minor reasons that also have significant impact in the performance of TCP links. One of the reasons for reducing the total bandwidth of the link can be the asymmetry of the link itself, this involves a disproportionate flow in one compared to the other direction. The fact that TCP uses the radio interface in both directions for confirmation of received segments is especially important not to over dimension the flow in one direction as opposed to the other.

Given that asymmetry is particularly expressed in the radio access technologies, it should always be kept in mind that the loss of packets or their delay will reduce the performance of the TCP flow. Therefore, there are certain conditions under which reduction of the performance of TCP should be taken into consideration:

- Acknowledgments of the received packets (ACK) may be lost or delayed due to the characteristics of wireless networks
- Speed of arrival of TCP packet confirmations determines the speed of sending of data packets. Such traffic consisting of confirmation of

packets is eruptive by its nature and the involvement of radio resources for each confirmation often leads to unwanted delays.

- ACK confirmations produce signaling in the lower levels that take a certain amount of resources in the opposite direction from the direction of generated data traffic.

TCP ideal situation is when the sender of the packets injects its segments in the TCP transmission pipe with the same speed with which the receiver takes them on the receiving side. The number of acknowledgments that are sent in the opposite direction is equal to the number of received segments by the receiver. Most important parameter upon which speed and quality of TCP links is measured is BDP (Bandwidth delay product) which represents a product of throughput and latency of the link. BDP parameter is of particular importance in the transport protocols based on a running window as TCP protocol. BDP value actually represents the amount of data transmitted from the transmitter, and has not been verified by the recipient and therefore defines the amount of data transmitter can send before the acknowledgment from the receiver that it has received the data. This also defines the minimum size of the transmit buffer in which to store transmitted and unconfirmed data, in case they need to perform their retransmission. BDP parameter is defined as:

$$\text{BDP } (\boldsymbol{bits}) = \text{Total\_awailable\_bandwith } (\boldsymbol{bits/sec}) \text{ x RTT } (\boldsymbol{sec})$$

Window for the TCP congestion control of the link in TCP links should be slightly larger than the BDP parameter and considering this fact the transmit window which is determined by the receiver should be greater than the BDP in order not to limit the congestion window.

Parameter related to packet delay marked as RTT, which is normally measured using the “Ping” application, is the most significant cause for performance degradation in TCP links. TCP uses a mechanism called "three way handshake" for implementation of reliability in transmission. TCP in the establishment phase of the link defines a typical window size of 1,5 x RTT. RTT time in wireless technologies is usually higher than the same in wired transmission technologies. Congestion control window during slow TCP start can be increased only in the time interval equal or greater to RTT time. This means that the greater the RTT time congestion control window will be increased with a slower speed. The longer the time of initial connection establishment and the longer it takes to reach the optimal window size, directly affects the reduction of the performance of the network and the flow throughput of data through a TCP link. It should be noted that the smaller the amount of data to be transferred, the impact is greater and performance is worse. For connections with larger theoretically achievable data flow certain rule applies: longer RTT time leads to less use of potential data flow as opposed to connections with smaller theoretical data flow.

II.9.4.3. UDP protocol performance in wireless networks

UDP protocol unlike TCP is not controlled and is unreliable protocol, which means that it does not guarantee the successful delivery of transmitted data. When a given network element needs to send a certain amount of information via UDP datagram, then it sends them without waiting for confirmation whether they have arrived at the opposite side. These features of UDP alone make them more resistant to changes in wireless environments, due to the simple reasons that there is no need for confirmation of arrived packets, there is no need for packet retransmission due to lost or erroneous packets, and there is no need for flow control and protection of piling.

For applications that require real-time communication, commonly used transmission protocol is UDP protocol, in which case they use medium or small datagrams. This implies that only a small part of the information is lost in the event of an error in the transmission of a datagram, but this configuration brings large overhead (due to the large number of datagrams that have appropriate header) in comparison with the situation when you would send smaller packets rather than larger datagrams that contain large amounts of information.

a. Application layer

Also there are degradations at the application layer. This means higher setup delay, however it makes it possible to give better QoS for the different flows. Application protocol logic is also something that has influence on the service performance. For instance, HTTP 1.1 delivers better performance than HTTP 1.0, because HTTP 1.1 uses the same TCP connection (and the same socket) for the main page (text html) and all linked objects to that web page, while the older HTTP 1.0 requires separate TCP connection (opening and closing sockets) for each individual object in the web page.

b. A new system for measuring quality parameters in wireless mobile networks

With the advent of GPRS as an upgrade of the existing GSM architecture, it paved the way for implementation of IP-based services within wireless mobile technologies. Integration of IP within the GSM architecture allowed the emergence of new services such as WAP (Wireless Application Protocol) and MMS (Multimedia Messaging Service), which represent a combination of services or service adaptation to user requirements and possibilities of the new architecture. Further development of GSM networks has give birth to EDGE (Enhanced Data Rates for GSM Evolution) as an upgrade in the radio network by adding more efficient modulation schemes for transmitting of data traffic in order to offer higher data rates to mobile users. The battle for higher data rates in mobile networks continued with the emergence of next generation 3G mobile networks and it still continues in order to meet the continuing customer needs for greater transmission speeds. Given that data services are aiming to become the leading source of revenue for mobile operators, the need for continuous measurement of their performance in terms of data quality of service offered to end users is of particular importance. For this purpose, the establishment of a system for qualitative testing of data services is of great importance. To claim that there is a system that performs proper measurement of the quality of the offered services, the claim must be based on real testing of appropriate key-performance indicators that provide qualitative parameters for each of the tested service. The group of these services includes popular data services in cellular wireless networks, such as: WAP, MMS, Web and E-mail. each of these data services depend on a number of transport layers on whom we can define different key performance indicators or KPI’s, and, as mentioned before, generally these indicators can be divided into three groups: basic indicators of network availability, network core sustainability indicators and core indicators of quality (packet delay, jitter, packet loss and speed of data stream or throughput). The goal, which appears as the subject of the measurement system for qualitative assessment of data service, is measurement of key performance indicators (KPI) associated to each data service. To meet this goal we made complete measurement architecture of measurement system (QPS - Quality Probing System) based on distributed measurement stations managed from a central location, which consists of the following system elements:

- QPS Server;
- QPS database;
- QPS user interface;
- QPS probes or measuring stations.

QPS system is designed to measure performance and analyse data services in GPRS / EDGE / UMTS mobile networks. Some of these services on which the performance measurements are conducted include all GSM data services such as SMS and CSD (Circuit Switched Data) via a standard dial-up connection, GPRS services such as MMS (Multimedia Messaging Service) and WAP (Wireless Application Protocol), and other IP-based services that use the same protocols over IP as well as over wired networks and are implemented through a GPRS connection.

Although all measurements within the QPS system are carried out by measurement stations, to be able to perform their measurement data processing, management of probes itself and their configuration, to define measurement scenarios, alarms in case of problems as well as graphics and numerical presentation of the measured results, all system elements of the QPS system listed above are required. QPS system architecture and logical connection between its elements are shown in Figure 26.

QPS system defines several so-called types of work procedures. The term working procedures within the QPS system include procedures that are different from other system procedures, which serve for the operational functioning of the system and use different delivery services (eg. USSD, CSD, SMS, etc.) on lower network levels according to OSI or use different upper layers according to the OSI (HTTP, FTP, Email, etc.).

The types of procedures are defined in Table 2.

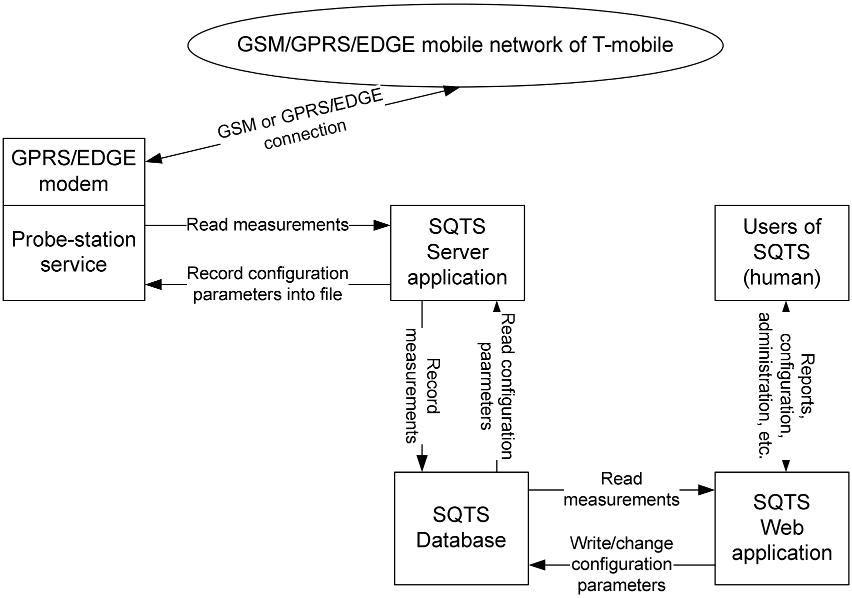

GSM/GPRS/EDGE mobile network of T-mobile
GSM or GPRS/EDGE connection
GPRS/EDGE modem
Probe-station service
Read measurements
Record configuration parameters into file
SQTS Server application
Users of SQTS (human)
Record measurements
Read configuration paarmeters
Reports, configuration, administration, etc.
SQTS Database
Read measurements
Write/change configuration parameters
SQTS Web application

Figure .26 Interactions between various elements of QPS system

Types of QPS work procedures
SMS (Short Message Service) procedure
MMS (Mobile Multimedia Service) procedure
WAP (Wireless Application Protocol) procedure
HTTP (Hyper-Text Transfer Protocol) procedure
FTP (File Transfer Protocol) procedure
SMTP (Simple Mail Transfer Protocol) procedure
IMAP (Internet Message Access Protocol) procedure
POP3 (Post Office Protocol version 3) procedure
Ping (i.e. ICMP – Internet Control Message Protocol) procedure
WS (Web Spider) procedure
WCS (Web Content Search) procedure
VAS (Value Added Services) procedure
WnW (Web and Walk) procedure

Table 2. Types of QPS work procedures

c. New concept of measuring the quality parameters in next generation wireless networks

Next generation networks consist of support functions for data transport, transport control functions, service control functions and applications and service support functions and applications. The first three functions are related to management of the performance of the networks in different ways. Functions for support of data transport consist of functions for network access, functions for core network support, gateway functions, and end functions. To provide support for the transmission of data information, as well as for the transmission of control and management information, performance measurement and management of these functions is needed. Transport control functions include access control functions and resource and admission control functions (RACF Resource and Admission Control Function) [40]. Resource and admission control functions to implement its core functionalities require reasonably accurate real-time data network resources and data on their real usage, and such information is necessary for the effective administration of the network resources and control decisions. The performance management control functions in next generation wireless networks can provide realistic view of information to the resource and admission control functions.

In addition to these measurements, particularly relevant are the performance measurements of control transport functions in order to portray realistic operation and control of data transport functions in next generation transport networks. Functions for control of the services and applications include resource control mechanisms, mechanisms for registration and authentication and authorization mechanisms on service level.

Measuring of traffic for control of the service is an essential activity in order to provide quality of service in the decision-making process during their establishment. This network entity is in constant interaction with various network architectural features of next-generation networks in order to collect and analyze performance measurements of the network and services. It can connect to the service control functions and applications for performance management of applications in next generation of networks as well as with service control functions. The results of these measurements can be delivered to the network entity for management of performance measurements.

II.10. Functional architecture for management of performance measurements

The functional architecture presented in Figure 27, is detailed overview of the general architecture of management of performance measurements in next generation networks.

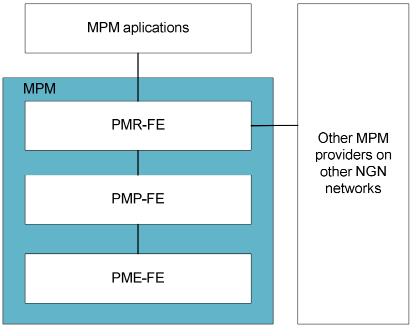


Figure 27. Functional architecture for management of performance measurements in NGN

The architecture consists of the following entities:

- *Performance measurement execution functional entity* (PME-FE) – presents an entity which measures performance. PME-FE is responsible for three groups of functions: performance measurement, processing of individual measurements and configuration enforcement entity. Execution of performance measurements includes initiation and termination of measurements on active probes same as passive measurements. Performing of individual measurements includes a collection of time stamped packets and calculation of their delay and loss in the individual probes. Configuration entity for conducting of measurements includes measurement configuration policies obtained from PME-FE.
- *Performance measurement processing functional entity* (PMP-FE) – presents an entity which processes the measurements. PMP-FE is responsible for two sets of functionalities: processing of measurement

results and configuration of the measurement probes throughout the network architecture. Processing functions of the measurements include the collection of measurement reports, their analysis, aggregation measurements and analysis of cyclical periods. This network entity actually entails individual results of performance measurements of PME-FE through the Mp interface and sends the results of its own analysis to PMR-FE network entity by Mr reference point (interface). Function for configuration of the measurement tests throughout the network architecture involves creating of the policies, selecting the reference measurement points where to set the configurations and the process of setting configurations in each measurement point.

- *Performance measurement reporting functional entity* (PMR-FE) – presents an entity for reporting of carried out performance measurements. PMR-FE performs collection of cyclic performance measurements of PMP-FE through the Mr interface and in the form of reports sends them to higher processing levels - MPM applications, such as RACF. This network entity also performs authentication of all requirements for initiating of measurements obtained via MPM applications and performs initiating of the requirements for performance measurements.

II.10.1. RTP / RTCP based-performance measurements

Most of the sessions of multimedia services over IP based networks are using RTP (Real-time Transport Protocol). RTCP (Real-time Transport Control Protocol) is accompanying the RTP protocol to transmit the feedback from receiver to sender’s RTP side. RTCP allows data recipients to perform proper estimation of data flow in direction towards them in a manner that perform computation of RTT towards the transmitters. Moreover RTCP-XR extended reports [b-IETF RFC 3611] provide a useful performance monitoring and diagnostics of VoIP service between the RTP transmitter and RTP receiver, such as user terminals (CPE's) in next-generation networks - NGN. In this context, several new types of blocks for measuring the performance of IP video services in terms of their implementation through the RTP/RTCP protocol were analyzed because of the very nature of the protocols - a firewall-friendly protocol. In addition to the RTP structure, the protocol defines RTP translator (intermediate forwarding service that performs RTP packet forwarding with their synchronization source without changing the source identifier) as a network entity with function for transmitting the measured performance information. For example, data flow corresponding to the data rate of TCP connections is periodically estimated using the values of the packet loss parameters and RTT time of the network which are obtained by using RTCP. Data headers of packets that are coming from users during normal traffic in the transport layer and lower layers do not contain time tags (timestamp) for the calculation of delay and variation in delay. These two parameters are very important for real-time applications such as voice and video. RTP protocol is an additional transport protocol to transfer packets from real-time applications, and as such is designed to be independent from the transmission network protocols. RTP packet in its header contains timestamp and sequence identifier, which allows performance measurement systems based on the RTP protocol to perform the correct calculation and estimation of packet delay variation of packet delay, if there is information about the application which communicates through each defined transport stream. In established RTP session besides presented measurement results there are also several other performance measurements sent in RTP packets that are very important in the process of quality assessment of the link such as: loss of data traffic or loss of fragments of data flow, as well as, inter arrival jitter. User measurement probes also evaluate the RTT time using this packets. With installation of relatively simple software agents in the user measurement systems based on RTP/RTCP, network providers can collect performance measurements without using additional active systems with external probes designed solely for performance measurement. The idea is to perform continuous performance measurements by the customer premises equipment (CPE) without introducing additional probe systems. Measurement results generated by this agent will be sent towards the performance management layer in order to provide aid in decision making process of resource allocation in next generation networks, and in that way directly affect the provisioning of the required QoS for each of the realized services that customers use. Figure 28 shows one possible system configuration for performance measurements using RTP/RTCP. In this figure separate Mu interface to Resource and Admission Control Function (RACF) is presented, which is an essential part of the performance decision making in the process of allocation of resources. By scaling this architecture for performance measurements to proposed NGN architecture in this book we can say that the interface for measurement data presented in this chapter is located between the QoS/QoE CM software module and software module for control of handovers between wireless technologies (ITHC).

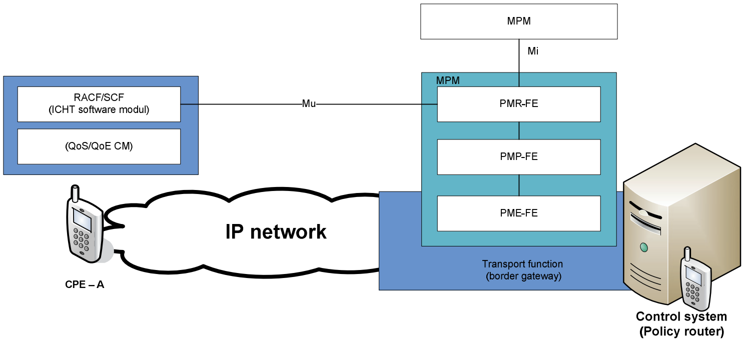


Figure 28. Network performance measurements using RTP/RTCP and RTCP extension.

Example of the implementation of the RTP/RTCP based control scheme based-performance measurements

Network performance between CPE's are measured and reported from MPM on a periodic basis, using the RTP/RTCP protocol using RTCP protocol extensions. In this example, MPM process the RTP/RTCP performance notifications as follows:

- CPE-A sends a request to the SCF for establishment of session or sessions with policy based router (QoS/QoE CM policy router) that plays a role of (CPE-B).
- RACF receives the request for the provision of resources for the service request along with other relevant information for incoming session or sessions of SCF by Rs reference interface which in our case is collocated with RACF; this information is sent to MPM via Mu interface.
- PME-FE MPM recognizes IP addresses and network ports to incoming RTP/RTCP sessions, after which it performs continuous monitoring of established sessions.
- PMR-FE receives the measured-performance information by PME-FE.

- Performance information report is generated in the PMR-FE and periodically sent to RACF through Mu interface.
- The report that is sent to RACF is often enriched with information about network resources, such as data flow, calculated by MPM.

The procedure for notifications based on qualitative performance measurements obtained by RTP/RTCP based scheme is shown in Figure 29.

The necessity of setting a new protocol for the exchange of network parameters between the two sides of the network architecture (policy client and router) is perceived by the parameters and scope of information to be exchanged.

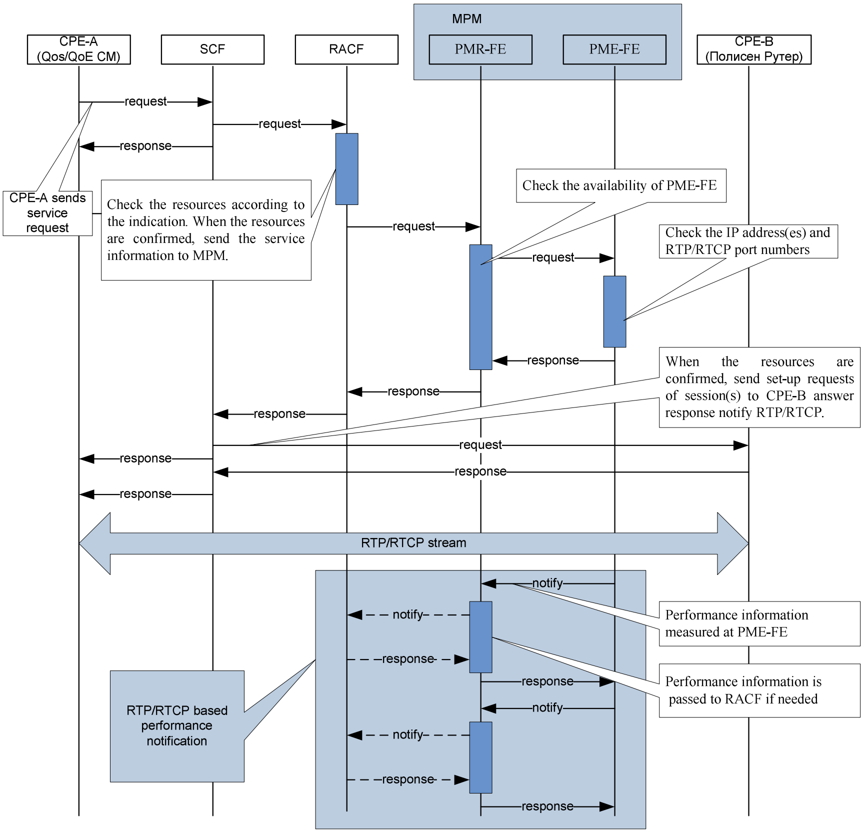

MPM
CPE-A (Qos/QoE CM)
SCF
RACF
PMR-FE
PME-FE
CPE-B (Полисен Рутер)
request
request
response
CPE-A sends service request
Check the resources according to the indication. When the resources are confirmed, send the service information to MPM.
Check the availability of PME-FE
request
request
Check the IP address(es) and RTP/RTCP port numbers
response
When the resources are confirmed, send set-up requests of session(s) to CPE-B answer response notify RTP/RTCP.
response
response
request
response
response
response
RTP/RTCP stream
notify
notify
Performance information measured at PME-FE
response
response
Performance information is passed to RACF if needed
notify
RTP/RTCP based performance notification
notify
response
response

Figure 29. RTP / RTCP-based diagram for performance measurements

Novel concept for performance measurement for the next generation of wireless networks

One of the basic activities of each network entity whose task is to control the quality of service is correctly identifying the requirements posed by each protocol level of the transmission network. That is to say, in the process of setting application service, depending on the type of application, each service has different requirements in terms of various quality parameters from the transmission network (capacity, delay, level of maximum permissible errors and lost packets during transmission, etc...). This establishment requires appropriate treatment of packets coming from different applications and is designated to specific Internet destinations. To be able to treat the packets received from the applications according to their required parameters, their recognition is a first step in the process of routing packets.

Recognizing the packets originating from a particular application with specific quality requirements is crucial in the process of policy routing, because these differentiated packets in the next stage should be placed and routed through heterogeneous network differently. In general there are various methods of determining the packets that belong to a given application, and they largely depend on the type of application and the extent to which the process of differentiation of the applications or services has to be broadened. We can say that good enough differentiation is the division of applications according to the protocol used by the transport layer (TCP, UDP, etc.) and corresponding port through which communication is established, which generally represents the definition of client socket that encompasses all the above parameters as well as the requested destination. In addition to this basic method there is an additional possibility to define application flow or a packet belonging to a particular application and that is based on deep packet inspection or analysis of packet. With the help of this analysis packets can be selected according to specific parameters within the application protocols themselves.

The purpose of the deep packet inspection is to perform separation of applications based on specific application parameters that are found within the application content of the packet. The best example of such an implementation is the separation of packets designated to various web destinations on different HTTP URL, which defines different applications.

Second step after packet differentiation that belongs to certain application is their appropriate routing through the heterogeneous network. As analyzed in the previous chapters, in the corresponding software module, adequate number of tunnel network interfaces are formed corresponding to the number of connected access wireless technologies by the client. The purpose of the routing module is to set policies for routing of the packets according to their background and their connection to higher level applications. These policies should include qualitative requirements according to the application and according to which appropriate selection of interface for routing of packets is made. Control of the routing is done by ITHC software module which using M-RATS algorithm selects the real-time transmission technology through which client transmits packets for a given application. Process of application initiation begins by defining its parameters in the process of creating a session within ITHC module as described under chapter II.3.

Defining and initiating the session depends on the parameters of the application. First step of this procedure is determining the parameters that characterize the application, and that are the basis for proper functioning of the packet detection process. As described above, there are various ways to describe a specific user application. One way is by describing the application parameters and by providing a way for their detection based on the presumed use of shallow or deep packet inspection. The definition of the application or its detection parameters described earlier can be filled in the system with manually mapping of each application in a predefined table of application parameters or they can be filled automatically during the process of installation of application. Such parameters define the application and shall be sent to the software module MQPBR in the process of initialization of the application.

Second step of the procedure for initiating a session represents the definition of the qualitative requirements requested by the application from the network. At this stage the qualitative requirements of the application, as described earlier, can be filled in to the system by manually mapping each application in a predefined table of qualitative parameters or automatically in the process of installation of application.

All of these parameters including definitions of application parameters and definition of qualitative parameters in the form of qualitative requirements by the network application and together with user defined preferences in terms of access technology preferred to serve specific application service form the so-called client application policy. Such client application policies exist for each application service which is accessed by the client, and depending on the implementation they can be automatically generated (based on certain predetermined parameters) or may suffer manual changes by the client, in case the client has different preferences and priorities defined in terms of information and importance of those applications.

Third step in establishing a session is the process of initiation of access network selection algorithm (M-RATS algorithm). Namely, it is necessary to create a separate instance of the M-RATS which will work for the particular application until the end of the user session for that specific application. This algorithm in accordance with the process defined in chapter II.4, performs continuous check/verification of the access technology states by examining the results of the M-RATS algorithm taking into account the qualitative requirements of the application and the real measured parameters of access technologies by QoS/QoE CM software module. The output of this process dictates the setting of rules (policies) for routing of the packets from a given application that is subject to analysis.

Last step in defining the session is the exchange of policies that define the parameters of the application and routing of found packets applicable to that defined application parameters. The exchange of these parameters would take place through a new protocol which as a basis will use DIAMETER protocol, over a special DIAMETER application designed for this purpose will be placed.

The need for setting up a new protocol for the exchange of network parameters between the two sides of the network architecture (client and policy router) is perceived from the volume of the parameters and information that should be shared.

For proper operation of the process for control and routing of packets, the exchanges of the following four key functionalities are required:

- Exchange of initial policies for technologies and user requirements;
- Exchange of the application characteristics for separation of the packets;
- Continuous periodic exchange of measurement information of the QoS parameters of each radio access technology;
- Exchange of information on selected set of routes for given applications (periodic exchange triggered by the ITHC module).

The idea of this protocol QoSPRO (Quality of Service Policy based Routing) is to unite all the essential and necessary parameters which by their exchange between the control and measurement entities, in particular between MQPBR and CQPBR software modules in the proposed architecture, would enable mutual synchronization in order to achieve proper operation.

Protocol procedures that should be supported by QoSPRO protocol are shown in Figure 30.

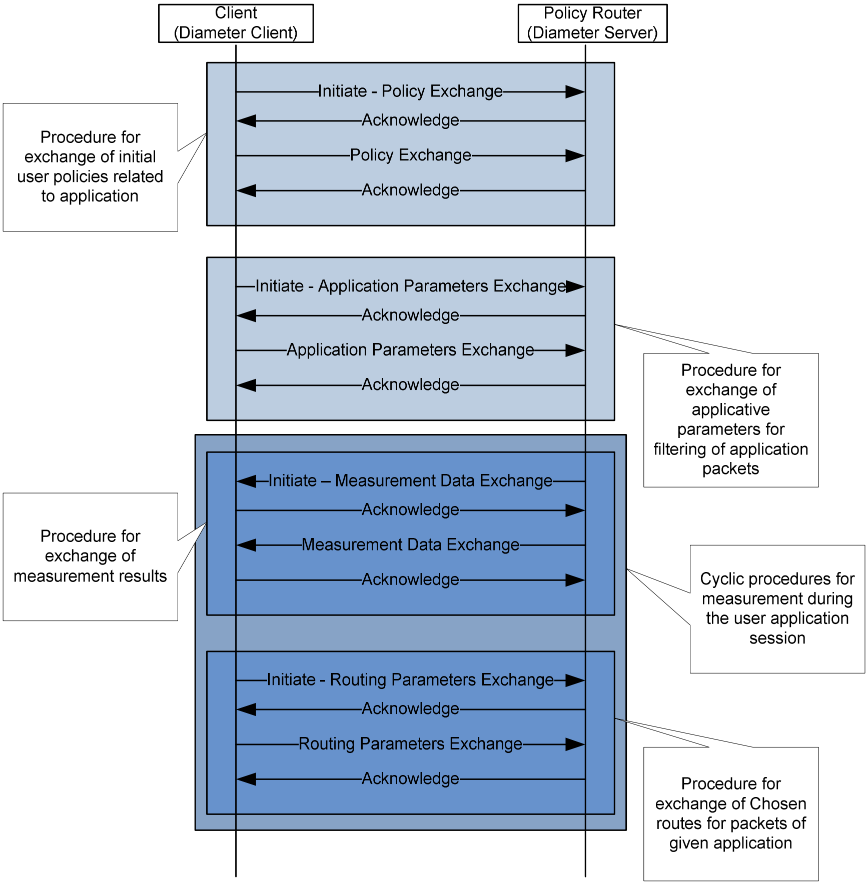

Client
(Diameter Client)
Policy Router
(Diameter Server)
Initiate - Policy Exchange
Acknowledge
Policy Exchange
Acknowledge
Procedure for exchange of initial user policies related to application
Initiate - Application Parameters Exchange
Acknowledge
Application Parameters Exchange
Acknowledge
Procedure for exchange of applicative parameters for filtering of application packets
Initiate – Measurement Data Exchange
Acknowledge
Measurement Data Exchange
Acknowledge
Procedure for exchange of measurement results
Cyclic procedures for measurement during the user application session
Initiate - Routing Parameters Exchange
Acknowledge
Routing Parameters Exchange
Acknowledge
Procedure for exchange of Chosen routes for packets of given application

Figure 30. **QoSPRO** – novel procedure for the exchange between network entities in the proposed architecture

Decisions for change of the radio access technology are performed on the mobile terminal (client) side, using the developed M-RATS algorithm [9], which uses key performance indicators as input parameters for proper decision for RAT choice.

All these information are used for proper direction of the packets of user applications in order to achieve best quality and optimal routing. It is also important to introduce hysteresis in the process of routing while performing periodic review of the status of the RATs, in order to avoid the effect of ping-pong switching between them. The idea for using the DIAMETER as the basis of this protocol gives great flexibility and simplicity and yet makes him a powerful and expandable enough in order to cope with all future requirements of the proposed architecture.

Conclusion

The development of the mobile and wireless networks is geared towards higher data rates and all-IP principle. Currently, there are many available radio access technologies which provide possibility for IP-based communication on the network layer; also it is noticeable that more and more services are migrating on IP environment including the traditional telephony and television, besides the traditional Internet services, such as web and electronic mail as most used among the others. On the other side, mobile terminals each year increase their mobile processing power; they have more memory on board, and longer battery life to serve the same applications (services). It is expected that the initial Internet philosophy of keeping the network as simple as possible, and giving more functionalities to the end nodes, will become reality in the future generation of mobile networks, here referred to as 5G.

In this chapter we have defined entirely new network architecture for such 5G mobile networks. The architecture includes introduction of software agents in the mobile terminal, which will be used for communication with newly defined nodes called Policy Routers, placed in the core network. The Policy Router creates IP tunnels with the mobile terminal via each of the interfaces to different RATs available to the terminal. Based on the given policies, the change of the RAT, i.e., vertical handover, is executed via tunnel change by the Policy Router, and such change is based on the given policies regarding the Quality of

Service and user preferences, as well as performance measurement obtained by the user equipment via new defined procedure called Quality of Service Policy based Routing (QoSPRO).

The proposed architecture for future 5G mobile networks can be implemented using common IP technologies (existing and standardized Internet technologies) with certain enrichment in their management and control, and its implementation is transparent to the radio access technologies, which makes it very likeable solution for the next generation of mobile and wireless networks.

Chapter III

EFFICIENT RADIO ACCESS TECHNOLOGY SELECTION FOR THE NEXT GENERATION WIRELESS NETWORKS

Introduction

Next generation of mobile networks will include many different wireless technologies, with different capabilities regarding the available bit rates, Quality of Service (QoS) support, mobility support, etc. In such heterogeneous environment it is very important to have intelligent mechanisms for access networks selection, as well as for periodical changing of the radio access technology upon given user constraints. The chosen access network at a given time period should be able to satisfy user requirements for a certain service, as well as to be supported by the wireless networks. Hence, radio network selection should incorporate characteristics of the network. For instance, IEEE 802.11 wireless networks do not have time division multiplexing and hence they cannot provide strict QoS support, while 3GPP mobile network (e.g., 3G, LTE etc.) and WiMAX can. However, intrinsic component for all wireless networks in the future is the IP on the network layer, which is the only possible internetworking technology, either as IPv4 or IPv6. Next generation mobile terminals should have capabilities for alternative network selection or simultaneous network selection [4].

One of the most important mechanisms for radio resources management in heterogeneous environment, in respect to the Quality of Service (QoS), radio resource utilization and user satisfaction, is the mechanism for initial network access for a given connection. This mechanism is referred to as initial choice of the radio access network and it consists of procedures for radio resource management for optimal choice of radio access network.

The main goal in this chapter is to analyze the new mechanism for access network selection, which will provide statistically better user satisfaction compared to other existing mechanism for such purpose. Better results mean better user satisfaction from given constraints regarding the velocity, type of service, required QoS and cost for service to the end user. Certainly, the demands from the network operators, service providers and users can be contradictory in certain scenarios. The proposed mechanism is based on usage of artificial intelligence for obtaining best user experience from the network selection in a heterogeneous wireless environment, based on measurements of different performance indicators taken by the mobile terminal or centralized network nodes. Such approach requires higher processing power for implementation of such algorithm on the mobile user side. However, with the increasing processing power and memory of mobile terminals, they will become capable of doing much more data storage and processing, and hence this foreseen development of the consumer equipment can be used for putting more intelligence in the radio access network selection. Such approach is outlined in this chapter.

Access network selection based on Fuzzy Logic and Genetic Algorithms has proven to provide better results and to be more robust when compared to random-based selection algorithms [13]. Furthermore, usage of nature inspired algorithms like Particle Swarm Optimization for optimization of Fuzzy Logic Controllers is analyzed in [14], [15], [16] and Multi-Criteria Decision Making systems are used in order to incorporate past knowledge of wireless networks behavior. All of these algorithms are generally based on learning capabilities provided from measured data, and therefore providing historical measurements of network related data is one of the main system elements, [17], [18], [19], and [20].

The main contribution of this chapter is in the development of an algorithm that enhances the way the Fuzzy Logic (FL) Controllers are build in a manner that optimizes FL decision, generated and optimized by Particle Swarm Optimization (PSO) algorithms.

Design of Fuzzy Logic Decision Algorithms

First part of the algorithm for selection of access network consists of array of fuzzy-logic controllers whose main task is to make fuzzification of the input variables in terms of analyzed criteria. Each of the fuzzy-logical controllers presents constructional unit of the joint parallel fuzzy logical system whose output values are inputs to/in the next stage of selection. This system is scalable and easily adaptable to changes in the heterogeneous environment. The system consists of four fuzzy-logical controllers: fuzzy-logical controller that refers to the level of reception of signal of an access network, fuzzy-logical controller which refers to the speed of mobile station, fuzzy-logical controller that relates to demanded Quality of Service by the user and fuzzy-logical controller that relates to the cost of the radio access technology.

Fuzzy-logical controller that refers to the level of reception signal of an access network as input variables takes the results obtained by measuring the reception signal level of access networks that are part of heterogeneous network, performed by the user at the point where it achieves the required service. The nature of such controller is to depict the network conditions that govern wireless networks. The low level of signal can cause unstable network connection and even complete network loss and hence we use fuzzy-logic for radio access selection, which is being fed from measurements made by the mobile device of the user. For analytical purposes, we use heterogeneous network consisted of two access networks WWAN (WCDMA) and WLAN (IEEE 802.11 b/g/n). As an input to the fuzzy-logical controller we have two variables for signal level, the first one SL1, which refers to WCDMA, and the second one SL2, which refers to the WLAN (Figure 31).

The range of values consists of all possible values according to guidelines and operational instructions of existing networks (WCDMA / UMTS, GSM) and (IEEE 802.11). Input variables SL1 and SL2 start from the lowest level that can be detected with/in the terminal (- 110 dBm) and end with the highest level of signal that can be received by the mobile terminal and is transmitted by the base station (-70 dBm and -50 dBm). Each value range is described by three so-called “linguistic variables”: low, medium, and high. Linguistic variable "low" assumes poor reception signal from radio access technology, the variable "high" indicates strong reception area of the receiving signal, while the variable "medium" indicates intermediate signal level. The shape of the membership functions in the space of values is presented in Figure 32 and Figure 33.

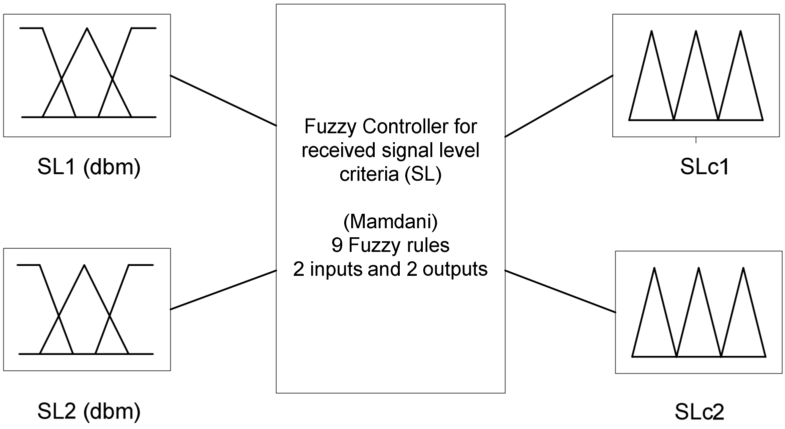


Figure 31. Fuzzy controller for received signal level from different wireless networks

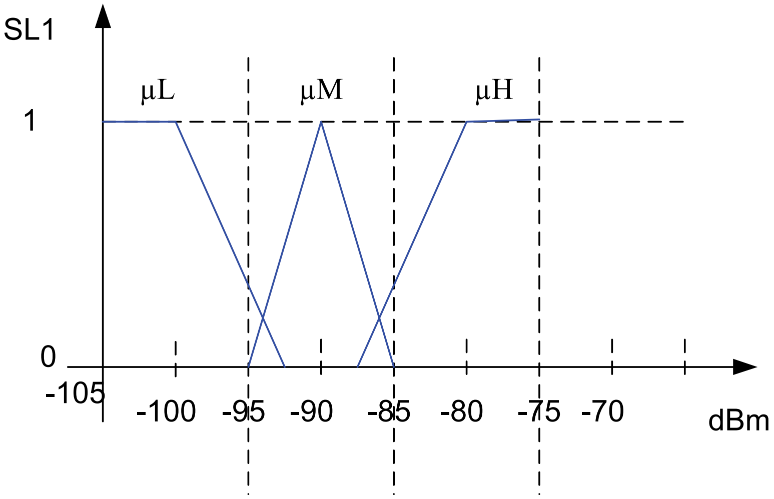


Figure 32. Membership functions for the variable shape of SL1

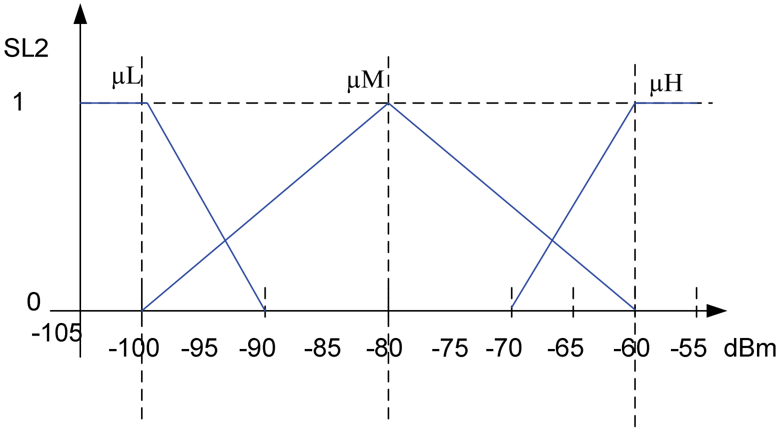


Figure 33. Membership functions for variable SL2

III.1. Fuzzy system of fuzzy associative matrices - fuzzy logic reasoning

Commonly used fuzzy systems in fuzzy-logic controllers are the Mamdani and Sugeno systems. Mamdani system is proposed in 1975 by Mamdani [21] and consists of four main parts. The main difference between the Mamdani system and other systems is that its usual output membership functions represent fuzzy sets. There is a defined set of stages for each variable to be defuzzified, after the process of aggregation,. Typical Mamdani rule system has the following form: If Input1=A and Input2=B then Output1= C and Output2 = D, where A, B, C, D are linguistic variables that describe the degree of membership of each variable. Sugeno or Takagi-Sugeno system is proposed in 1985 [22]. It is quite similar to the Mamdani system considering many similar features, however its main difference is that the output membership functions in Sugeno systems are linear functions or constants. As a result of this there is no need for defuzzification process. The result is a fixed constant value. Usually the rule in Sugeno systems has the following form: If Input1 = X and Input1 = Y, then output Z = a*X + b*Y + c. In Sugeno system of order zero, output level z is a constant, z = c (a = b = 0). In these systems output functions are set using a single value, called "singleton" sets. These sets have a single value "c" into a single point of value range as having zero value in all the rest. Its implementation is simpler than the Mamdani system that requires significant computational effort and resources, but on the other hand, they don't provide easy way of showing the inherent linguistic rules of human reasoning. On the contrary, in Mamdani system total flexibility is presented in the way of choosing fuzzifier, a system of reasoning, and defuzzification, making it effective in the manner of representation of fuzzy logic close to humans and therefore represents the first choice for fuzzy system analyzed in this chapter. Review of the rules is presented in Figure 34.

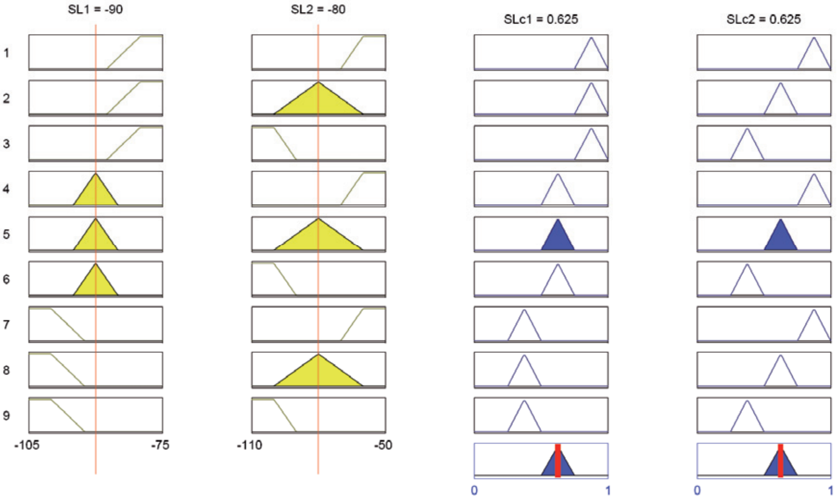


Figure 34. Review of rules for reasoning in fuzzy-logic controller for level entrance signal

Considering that for fuzzy-logic controller we have chosen a Mamdani approach, output variables are fuzzy sets defined by fuzzy variables. Regarding this defuzzification, for each output there is a set of four triangular membership functions that represent the four linguistic variables (not acceptable, probably not acceptable, probably acceptable, and acceptable). Defuzzification process is the same for both output variables and its appearance can be described with Figure 35.

Each fuzzy-logical system has two output variables that describe the likelihood of user to be assigned to one of two proposed access network technologies. The design of a fuzzy-logical system in general uses five different methods for defuzzification, namely: the method of Centroid of Area - CoA, the method of Bisector of Area - BoA, the method of averaging the maximum (Mean of Maximum - MOM), the method of Smallest of Maximum - SOM, and the method of Largest of Maximum - LOM. During the evaluation, the method of CoA will be used, as it has several advantages over the other methods. It is the most accurate method and it is commonly applied. At the same time, it did not neglect the forms of output membership functions as methods of maximization and does not suffer from ambiguity problems, because there are precisely defined conditions. On the other hand, it has great complexities, which in the proposed algorithm are greatly reduced because of the use of simple output membership functions such as triangle functions. Logical predicates are represented in Table 3 in the form of IF, THEN rules.

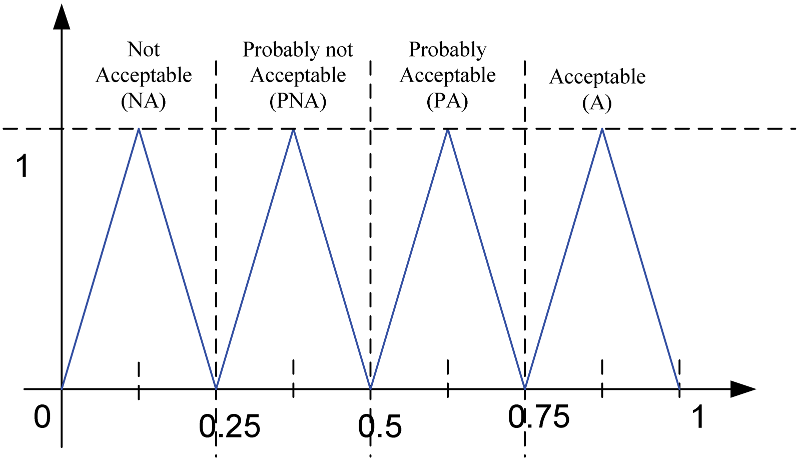


Figure 35. View of defuzzifier for SLc1 and SLc2 variables

Nr. of the rule	**Rule definition**
1	If SL1 is high and SL2 is high then SLc1 is acceptable and SLc2 is acceptable
2	If SL1 is high and SL2 is medium then SLc1 is acceptable and SLc2 is probably acceptable
3	If SL1 is high and SL2 is low then SLc1 is acceptable and SLc2 is probably not acceptable
4	If SL1 is medium and SL2 is high then SLc1 is probably acceptable and SLc2 is acceptable
5	If SL1 is medium and SL2 is medium then SLc1 is probably acceptable and SLc2 is probably not acceptable
6	If SL1 is medium and SL2 is low then SLc1 is probably acceptable and SLc2 is probably not acceptable
7	If SL1 is low and SL2 is high then SLc1 is probably not acceptable and SLc2 is acceptable
8	If SL1 is low and SL2 is medium then SLc1 is probably not acceptable and SLc2 is probably acceptable
9	If SL1 is low and SL2 is low then SLc1 is probably not acceptable and SLc2 is acceptable

Table 3. Fuzzy rules for decision-making for level signal reception

Fuzzy-logic rules are established to assign the user to a network that has the best radio performance and better level of reception signal. Such implementation would provide a stable system with a reduced number of handovers due to low levels of signal reception and good conditions for the initiation of service. This principle of creation of fuzzy-logic rules apply to all fuzzy-logic controllers in the system for access network selection.

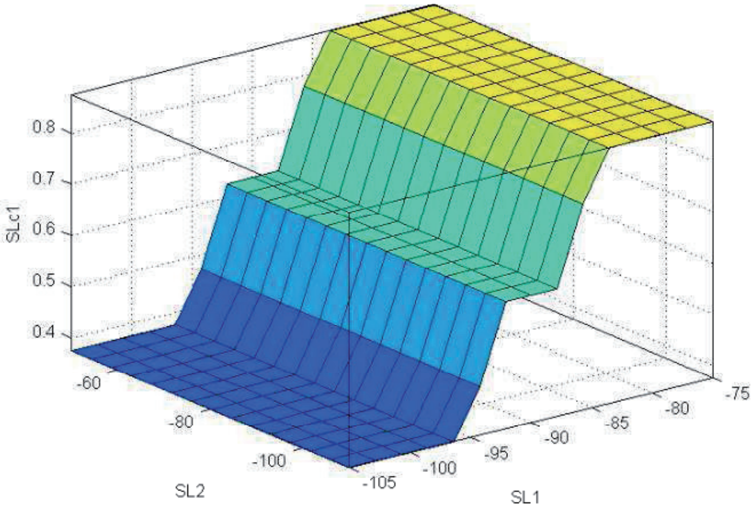


Figure 36. Control input-output plane of the output variable SLc1

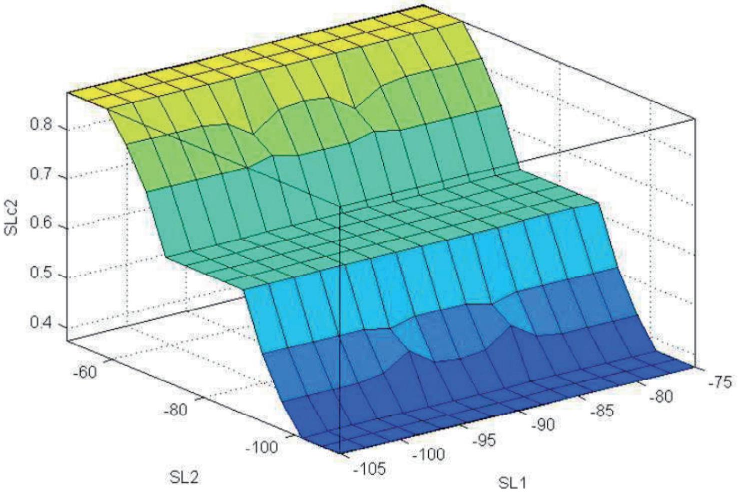


Figure 37. Control input-output plane of the output variable SLc2

III.1.1. Input-output Fuzzy control surface

Nonlinear mapping of the input variables into the output variables implemented in some fuzzy-logic system is called input-output fuzzy control surface. Mapping done in this way can be displayed as non-linear surface that represents all phases of the information system in a compact way. The surface is created with implementation of fuzzy rules and membership functions. In its creation we have made interpolation between the rules of fuzzy-logical system. The output represents an interpolation of the effects of the rules that are activated at a given time. Input-output Fuzzy control surfaces for signal level of the mobile network (WWAN) and WLAN are shown in Figure 6 and Figure 7. It can be seen that by increasing the value of SL1 the probability to choose a WWAN network is higher (the value of SLc1 is greater), while on the other hand, increasing the value of SL2 the probability to choose a WLAN network is higher, which is expected behavior of the proposed fuzzy-logic system.

III.1.2. Fuzzy-Logic Controller for the Speed of the Mobile Terminal

Fuzzy-logic controller which refers to the speed of the mobile terminal receives as an input variable results obtained by measuring of the speed of the terminal, carried by the user at the point where it achieves the required service. The criterion for the terminal velocity reflects the requirements of the operator within the algorithm for selecting the access network. Connecting of the slow moving users to access networks with less coverage, such as WLAN networks, and users who move with higher speeds to networks with broad coverage reduces the number of unnecessary handovers and conserve the resources of the heterogeneous network. Design of the fuzzy-logic controller is in direction of providing the above requirements. The system has one input, the terminal velocity (TV), and two outputs (TVc1, TVc2) related to the values of access networks considered in the system (WWAN and WLAN), as illustrated in Figure 38. The range of values for speed of the terminal is set between the resting position of the user (0 km/ h) and the position of running (10 km/h). All other values above this are taken as values with a high level of mobility of the terminal. In terms of linguistic variables in the value range, three values that represent the mobility of terminals are defined as: low, medium and highly portable mobile terminals (Figure 39).

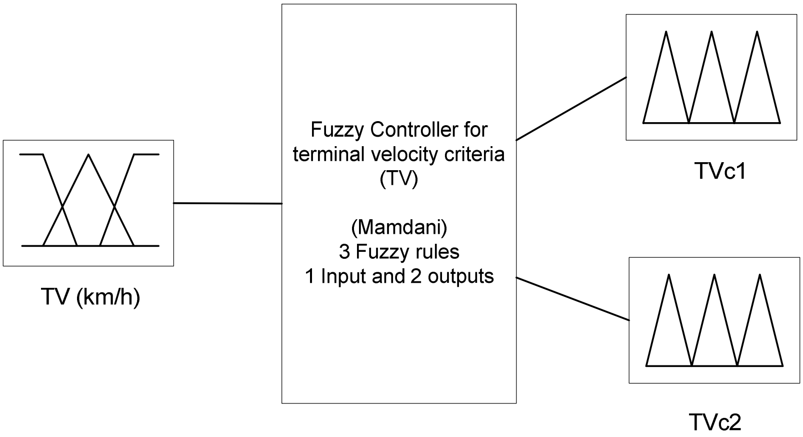


Figure 38. Fuzzy controller for the speed of mobile terminal

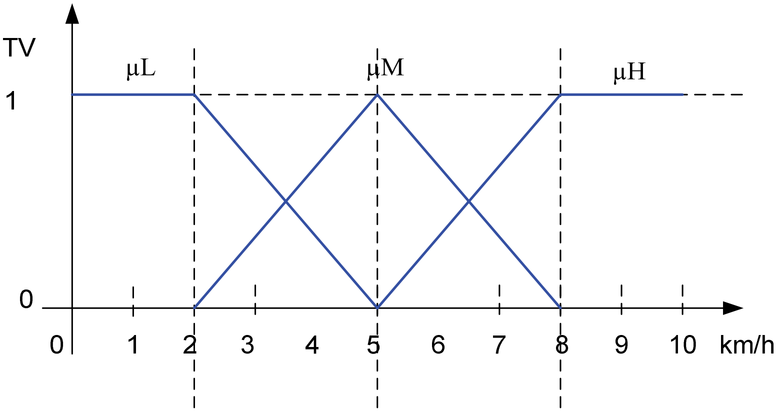


Figure 39. Membership functions for variable speed of the mobile terminal

Nr. of the rule	Rule definition
1	IF TV is low Then TVc$_1$ is not acceptable and TVc$_2$ is acceptable
2	IF TV is medium Then TVc$_1$ is probably acceptable and TVc$_2$ is probably acceptable
3	IF TV is high Then TVc$_1$ is acceptable and TVc$_2$ is not acceptable

Table 4. Fuzzy rules for terminal velocity for Decision Making in Fuzzy Systems

Fuzzy-logic controller which refers to the speed of the mobile terminal has three fuzzy rules. The system is designed to minimize the number of handovers within the heterogeneous wireless network and consequently to increase usability of network resources. In the form of IF, THEN rules logical predicates are represented in Table 4.

As it can be seen from the graphs of input-output control surface, with increasing speed of the terminal, output value for WWAN access network TVc1 smoothly increases towards full acceptance and choice of this network as the best alternative (Figure 40), while the output value for the WLAN access network TVc2 smoothly decreases towards complete rejection (Figure 41).

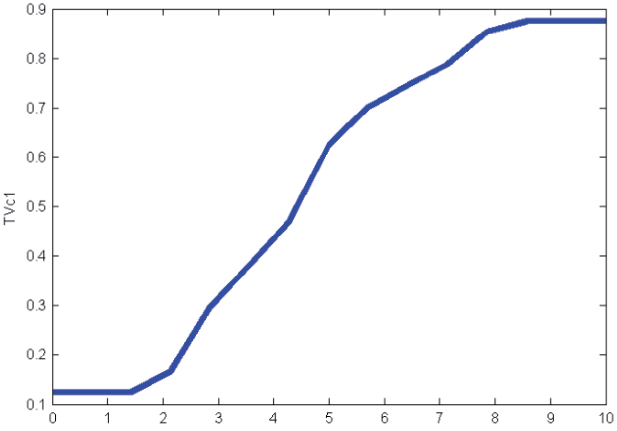


Figure 40. Control input-output plane of the output variable TVc1

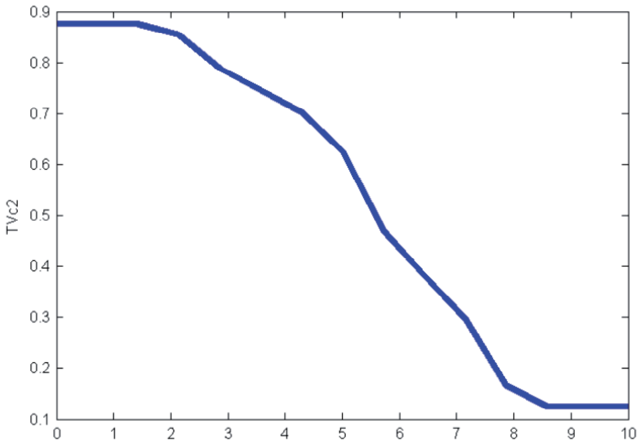


Figure 41. Control input-output plane of the output variable TVc2

III.1.3. Fuzzy-Logic Controller for Quality of Service

Fuzzy-logic controller related to the Quality of Service, required to be meet by the mobile terminal, has as input variables requested packet delay and packet speed (bit rate), which are required by the service for its proper functioning. The criterion for Quality of Service reflects the requirements of services and applications within the algorithm for access network selection. Fuzzy-logical system has two inputs, requirement for maximum delay of packets (TS1) and requirement for minimum bit rate (TS2) as independent inputs, and two outputs (TSc1, TSc2) relating to the values of access networks considered in the system (WWAN and WLAN), as shown in Figure 42. Input variable that presents the request for the maximum delay of packets, describes the required maximum delay of packets from end to end through the system required for proper operation of the service, while the request for minimum bit rate of the access network is to accommodate for minimum bandwidth for providing the service. Values of the two input variables are determined taking into account the most sensitive services for each variable (e.g., voice call is most sensitive application in terms of time delays), and insensitive services for each input variable (e.g. packet non real time services are examples of services with the least rigorous requirements on the basis of time delay of packets). Every value space is defined by three linguistic variables (high, medium, low). Value range of default package starts with delays between 0 and 200ms (milliseconds), which is typical for conventional voice services in real time, and ends with a delay of more than 800ms (milliseconds) which is appropriate for background non-real time services (Figure 43). Value range for the required bit rate (Figure 44) starts with a request for bit rate of less than 50kb/s, which is typical for services with low bit rate, services like voice or video services in low resolution, and ends with the bit rate of more than 350kb/s that is typical of packet based traffic with high speed data transfer or video traffic with high resolution (for a mobile terminal).

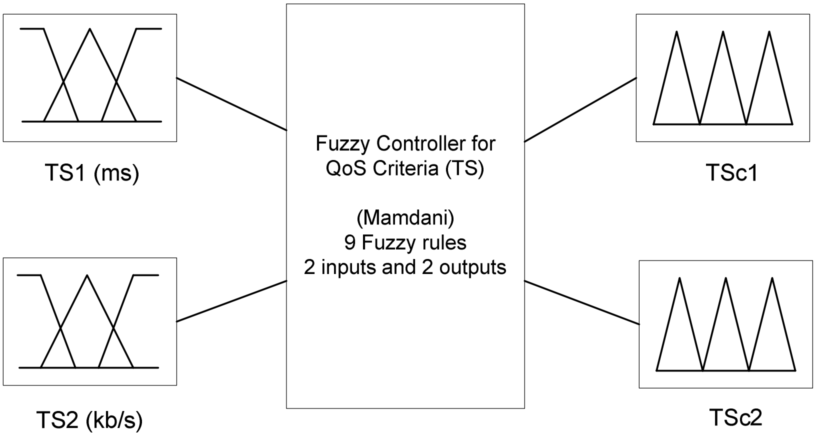


Figure 42. Fuzzy Controller - Requirements for Quality of Service

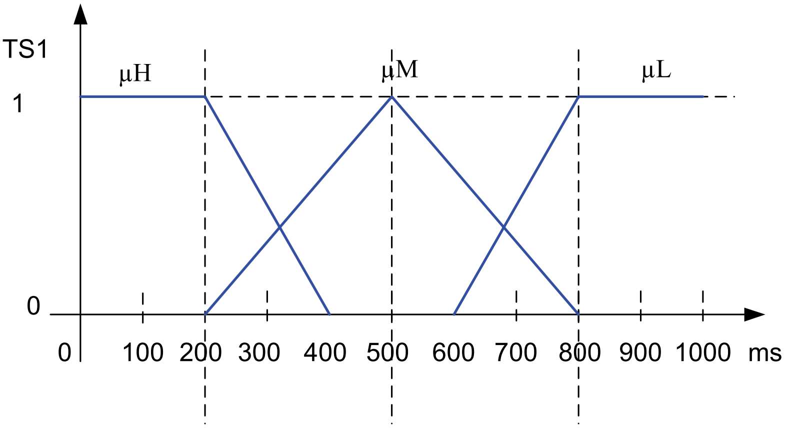


Figure 43. Membership functions for the required time delay of packets

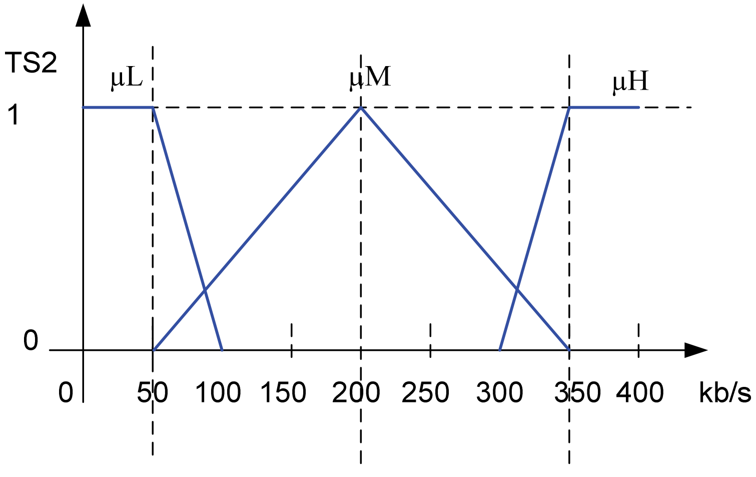


Figure 44. Membership functions for the required bit rate

Fuzzy-logic system consists of 9 fuzzy rules. Each rule is set in order to satisfy the application and service requirements of the system for assigning stages. As a result of this, the rule matrix for decision making is defined in Table 5.

Nr. of the rule	Rule definition
1	If TS1 is high and TS2 is high then TSc1 is acceptable and TSc2 is probably not acceptable
2	If TS1 is high and TS2 is medium then TSc1 is probably acceptable and TSc2 is probably not acceptable
3	If TS1 is high and TS2 is low then TSc1 is probably acceptable and TSc2 is probably acceptable
4	If TS1 is medium and TS2 is high then TSc1 is probably acceptable and TSc2 is probably not acceptable
5	If TS1 is medium and TS2 is medium then TSc1 is probably acceptable and TSc2 is probably not acceptable
6	If TS1 is medium and TS2 is low then TSc1 is probably not acceptable and TSc2 is probably acceptable
7	If TS1 is low and TS2 is high then TSc1 is probably not acceptable and TSc2 is probably acceptable
8	If TS1 is low and TS2 is medium then TSc1 is probably not acceptable and TSc2 is acceptable
9	If TS1 is low and TS2 is low then TSc1 is not acceptable and TSc2 is acceptable

Table 5. Fuzzy rules for deciding the type of service in fuzzy system

Service Assignment of applications with stringent requirements for time delay is done on a WWAN access network and service requirements for high bit rates are assigned to the WLAN access networks. Other rules follow the direction of this value in both ranges. To highlight the benefits of the operator, rules can be modified in the direction of providing more service types to WLAN networks while freeing up more resources in the WWAN network for serving the premium services.

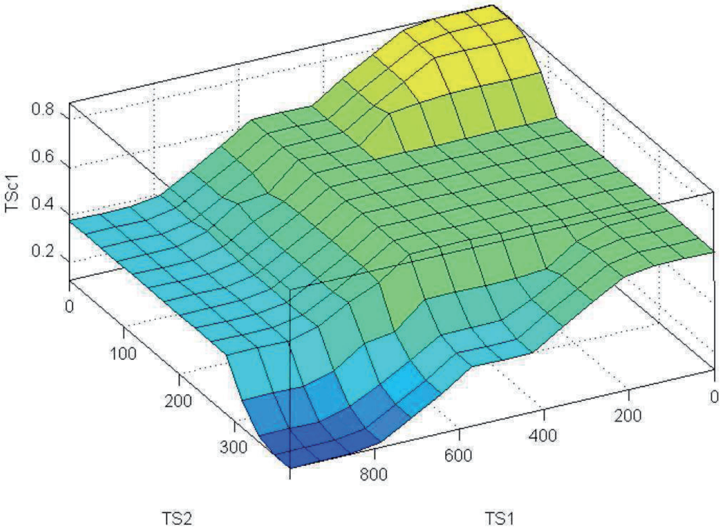


Figure 45. Control input-output plane of the output variable TSc1

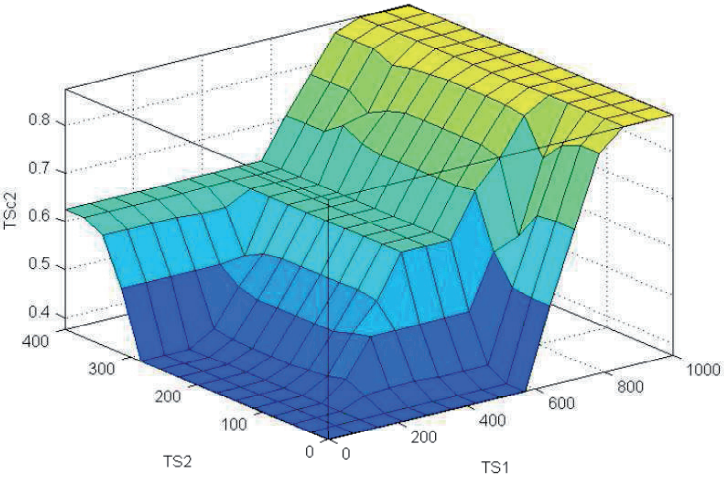


Figure 46. Control input-output plane of the output variable TSc2

The control input-output area for variables TSc1 and TSc2 represents the dependencies of output variables (that represent the choices of an access network) upon input parameters which are characterized by the type of service, delay of packets and required bit rate. By presenting surfaces for both output variables in Figure 45, one can conclude that, by reducing the value of time delay (delay is the more stringent of the requirements) or by reducing the required bit rate, probability of selection of WWAN access networks is higher. The same applies if the two variables (delay and bit rate requirements) are reduced simultaneously. On the other hand, from Figure 45 can be concluded that by increasing the amount of packet delay or increasing the required bit rate, probability of selection of the WLAN as access network is higher. The same applies if the two variables (time delay and required bit rate) increase simultaneously.

III.1.4. Fuzzy-logic Controller for Cost of the Radio Access Technologies

The Fuzzy-logic controller referred to as cost of access technologies, as an input variable takes into account the user-defined cost for each access technology that builds the heterogeneous network. The criterion for the cost of access technology reflects the user requirements within the algorithm for selecting the access network. The purpose of this fuzzy-logic controller is to make connection of the users with given access technologies in terms of cost, taking into consideration the fact that users who want a lower cost of service are primarily directed toward low-cost technologies such as WLAN, and users who want the service without compromises in terms of cost are turned to more expensive networks like WWAN.

Design of fuzzy-logic controller is conducted in a way to provide the above mentioned requirements (Figure 47). The system has one input, and it uses requested cost (PR) as an input, and two outputs (PRc1, PRc2) related to the values of access networks considered in the system (WWAN and WLAN). Value range of the input variable is selected in a way that user costs are scaled in the range of 1 to 10, which covers all possible values between the position of the free service (value 0) and the maximum possible price (value 10). In terms of linguistic variables in range of values three values are defined that represent the cost of technologies, these are: low cost, medium cost and high cost (Figure 48).

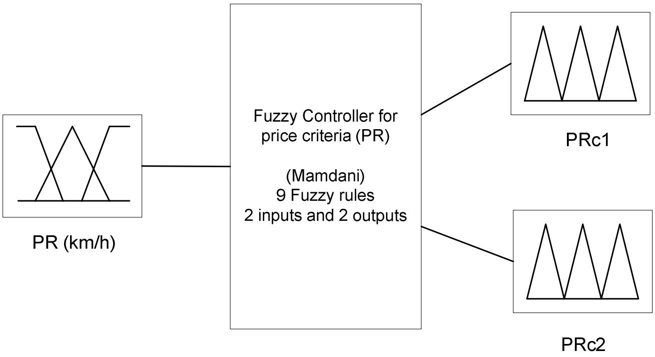


Figure 47. Fuzzy controller for the cost of radio access technology

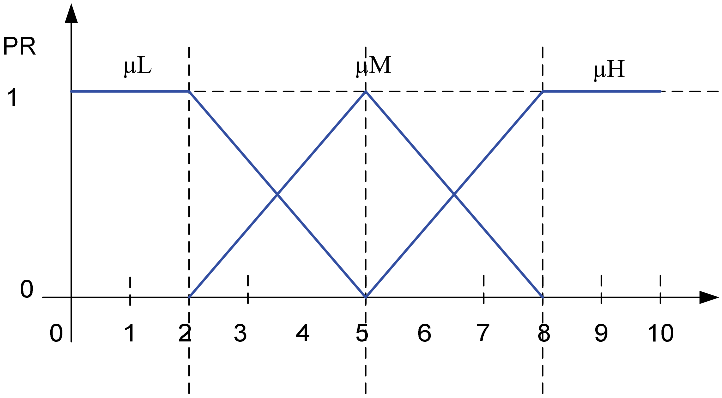


Figure 48. Membership functions for the variable, cost of access technology

Nr. of rule	Rule definition
1	IF PR is low Then PRc_1 is not acceptable and PRc_2 is acceptable
2	IF PR is medium Then PRc_1 is probably acceptable and PRc_2 is probably acceptable
3	IF PR is high Then PRc_1 is acceptable and PRc_2 is not acceptable

Table 6. Fuzzy rules for cost of the radio access technologies

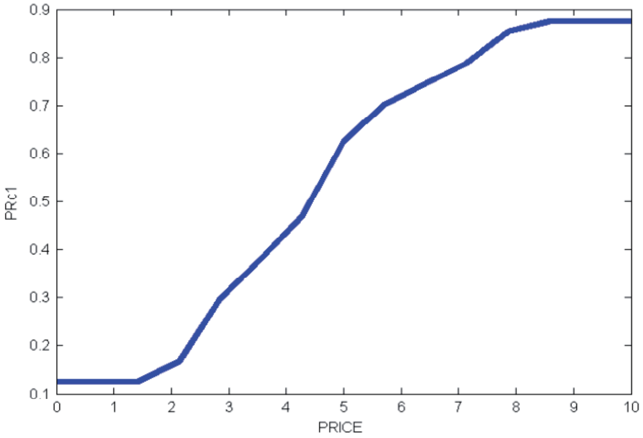


Figure 49. Control input-output plane of the output variable PRc1

Fuzzy-logic controller that refers to the cost of access technology has three fuzzy rules. The system is made in a way that users with demands for cheaper service are directed towards WLAN access technologies while other users are using WWAN technologies. In the form of IF, THEN rules, logical predicates are represented in Table 6.

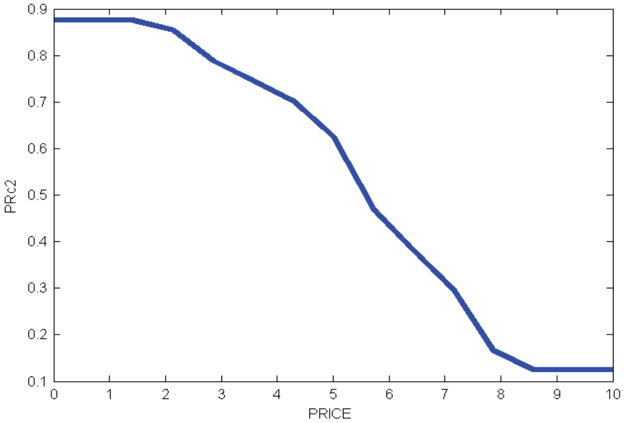


Figure 50. Control input-output plane of the output variable PRc2

As shown in Figure 49, in input-output control area by increasing the required cost the output value PRc1 for WWAN access network controller smoothly increases towards full acceptance and choice of this network as the best alternative. In contrast, output value for the WLAN access network PRc2 (Figure 50) the controller smoothly decreases towards complete rejection. Such behavior is in full accordance with the input requirements for the design of this controller.

III.2. Fundamentals of particle swarm optimization algorithm (particle swarm optimization)

Algorithm based on particle swarm optimization (PSO) is a stochastic optimization technique implemented on a given population of possible solutions inspired by social behavior of a particular group of animals such as birds or fish. The algorithm uses a population of particles that form a cluster that moves through the space of possible solutions. The system is initialized with a set of randomly generated population of possible solutions. Then, the algorithm goes through the process of optimization where it searches the optimal solution with periodic checks on new generations obtained by advancement in the solutions made by the whole population. PSO potential solutions, called particles, move through the space of possible solutions to the problem by following the current optimum particles made in the observed phase of the algorithm. In this operation each particle regulates or models their position under its acquired experience during the movement and according to the experience of its neighbors, companions. Each particle keeps track of its position in the space of possible solutions and is treated as a point in N-dimensional space of solutions which is directly connected with the best value match with a possible solution to the moment of observation. This value is called "pbest" or the best value of the particle. Despite this value optimizer keeps a record of all the times the best value is achieved by all particles of the population and it is called "gbest". This value represents the global best value and its final value with termination of optimization process presents an ultimate solution from optimization process. The basic concept of PSO is based on introduction of random acceleration in the movement of each particle in the direction of the position of his "pbest" and "gbest" as shown in Figure 51. Namely, each particle "i" in the course of their movement remembers its best position in the space of possible solutions "pbesti", and at the same time remembers the best position of its neighbors from the topologically closest neighborhood. Depending on the way you define the area of the neighborhood we can look at two different values "lbest" if it is a local version of the algorithm or "gbest" if the algorithm covers the entire population as the neighborhood of the particles. PSO algorithm is an iterative evolutionary algorithm. In each iteration particle "i" adjusts its speed vij and position ***pij*** through each dimension "j", in terms of its personal best position "***pbestij***" and the best position of the whole swarm (flock) "***gbestj***", if used global variant algorithm where the neighborhood of each particle is covered by the entire population. Adjustment is carried out through the following expressions:

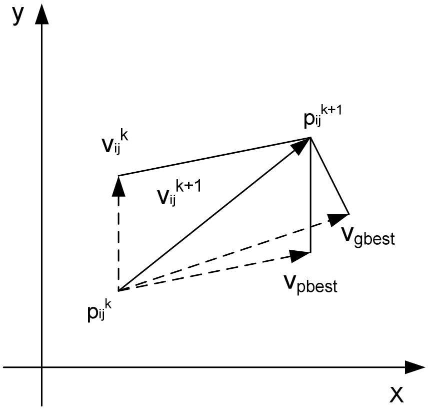
$$v_{ij} = k(v_{ij} + c_1 r_1 (pbest_{ij} - p_{ij}) + c_2 r_2 (gbest - p_{ij})) \tag{1}$$

$$p_{ij} = p_{ij} + v_{ij} \ , \tag{2}$$

where c1 and c2 are constants of acceleration, while r1 and r2 are random real numbers from the set U (0, 1), and k represents the limiting weight factor. According to M. Clerc, J. Kennedy [16], [17] it can be found that in optimization algorithms based on particle swarm is necessary to introduce a limiting weight factor k. to ensure their convergence. This factor is defined as:

$$k = \frac{2}{\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|} \tag{3}$$

where φ = c1 + c2. The introduction of such a restriction allows the movement of particles through space solutions to positions "***pbestij***" and "***gbestj***" to be done on navigated way, leaving the freedom of each particle space to explore new potential solutions in the neighborhood, moving through with a random direction and speed, avoiding the local minimum, as shown in Figure 51. This algorithm terminates when it reaches the maximum number of iterations or after a sufficiently large number of iterations the best position of the entire swarm cannot be improved.



p_{ij}^{k} : current position of the particle
p_{ij}^{k+1} modified position of the particle
v_{ij}^{k} : current speed
v_{ij}^{k+1}: modified speed
v_{pbest} : speed based on ***pbest***
v_{gbest} : speed based on ***gbest***

Figure 51. Concept modification of the position of the particle in the process of searching for by using particle swarm optimization

III.2.1. Use of the algorithm based on particle swarm optimization for defining of fazilogic controllers

System based on particle swarm optimization is set together with each fazilogic controller and its purpose is to carry out the optimization of certain elements of fazilogic system. In principle it is possible to complete optimization of the fazilogic controller on each of its building blocks, but due to the simplicity of the system it is best to perform optimization only of a segment of fazilogic controller and that is the domain (range) of values of each of the membership functions within the fazificator. Optimization method relies on human knowledge and subjective scores for correct and proper selection of the solution to point the way of the optimization algorithm. Moreover, the guidance of the algorithm does not define the ultimate solution, but the direction towards the desired goal, and its optimization features are expected to find the best solution that meets the set norms. Optimization process begins by generating an initial population (cluster) of real values (particles) that are randomly generated in the range defined by the membership functions or their domain values. Each particle consists of randomly generated limit values of the membership functions. After the process of generating initial population the cycle of optimization starts. The first step in this cycle is the process of restricting the values of particles which is done by applying limiting function that defines the minimum and maximum limit value of each particle and prevents overflow values and their disruption of the function of belonging. The second step includes movement of particles through space of values provided by specified function of the movement. After each movement rechecking of limitations on new acquired values is performed. Consequently, by applying the evaluation function on cluster particles the value of the degree of acceptability for each particle is determined using MSE (Mean Square Error) calculation between the values of the newly formed particle and values assumed best original particle obtained by human knowledge from established fazilogic controller. Such defined errors are collected and they determine the target maximum error of the whole flock. Knowing the best particles the others begin the process of moving analogous to the collection of birds in the flock. Upon completion of this process, the cycle is repeated until obtaining the best value determined or until the expiration of the anticipated generation, depending on the fact which constraints will be reached first.

Considering the algorithm itself and after experimental optimization as best values in the evaluation process we can define population of 100 particles and a maximum number of 50 iterations. As an ending constraint regarding maximum error value for all particles in the population is defined by limit of 20. In the same way parameters in the movement process are defined as acceleration constants: cognitive acceleration c1 = 2,5 and social acceleration c2 = 2,7 acceleration neighborhood is not used and the corresponding parameter has a value of c3 = 1. After the optimization process is finished optimized fuzzy-controllers are obtained, whose membership functions best reflect current conditions in the system of access technologies that build the heterogeneous network.

III.2.2. Multi Criteria Decision Making Algorithm for Selection of Radio Access Network

Main purpose of the multi criteria decision making system is to perform appropriate classification of the considered alternatives according to their acceptability. It strives to provide: highest level of satisfied customers, more users that have best quality of service, conservation of resources of the networks with a higher cost in a way that utilizes the networks with less cost. Having in mind the main purpose for using the systems for multi criteria decision making, two possible alternatives are considered as technologies for the selection of WWAN and WLAN access network. As input into multi criteria decision making system generally outputs derived from fuzzy-logic controllers analyzed in the previous chapters are used and are obtained as a result of the processing of following criteria: level of reception signal, speed of the terminal, type or Quality of Service and cost. In the process of decision making algorithm we use an algorithm for evaluation of multiple attributes. This algorithm is linear method for making decisions, which makes it easy to use in hybrid and very complex models. Use of fuzzy logic with PSO optimized fuzzy controllers in terms of optimization of input variables and genetic algorithms (GA) in order to adjust internal processes of the algorithm for evaluation of multiple attributes has all dynamic advantages from artificial intelligence to meet the specific requirements involved in the process of selecting radio access network. Input of this algorithm are results obtained from the evaluation of input variables through fuzzy-logic systems for given input criteria: (SLc1, SLc2; TVc1, TVc2; TSc1, TSc2; PRc1, PRc2). Given that all outputs from fuzzy-logic systems are in the range of [0, 1] scaling of variables is not needed. The input criteria form matrix A given as:

$$A = \begin{pmatrix} SLc_1 & SLc_2 \\ TVc_1 & TVc_2 \\ STc_1 & STc_2 \\ PRc_1 & PRc_2 \end{pmatrix} \tag{4}$$

The matrix of weight factors CF by given criteria is presented as:

$$CF = \begin{pmatrix} F_s & F_v & F_q & F_p \end{pmatrix} \tag{5}$$

where F_s is the assigned weight factor for the criterion based on the level of reception signal, F_v is the weight factor for the criterion based on the speed of the user, F_q is the weight factor for the criteria by type of requested service and F_p is the weight factor by the criterion of cost. The value of the weight factors are positive numbers. If they are real numbers then they should be within the limits of [0, 1] and should satisfy the requirement that their sum is equal to 1.

$$F_s + F_v + F_q + F_p = 1 \tag{6}$$

If they are positive integers then they should be within the range of [0, 100], and their sum must be equal to 100:

$$F_s + F_v + F_q + F_p = 100 \tag{7}$$

The values of weight factors can be assigned manually according to the experience of decision makers and according to its knowledge of the weight of each criterion on the selection process of the radio access network, or using a particular method of optimization, such as genetic algorithms, where their value is obtained through the process of moving in the genetic optimization algorithm towards pre-specified goal. Ranked values or outputs of multi-criteria decision making algorithm for deciding on alternatives are presented as: X_{WWAN} and X_{WLAN} and are calculated using the following equations:

$$X_{WWAN} = \frac{SLc_1 * F_s + TVc_1 * F_v + STc_1 * F_q + PRc_1 * F_p}{SF} \tag{8}$$

$$X_{WLAN} = \frac{SLc_2 * F_s + TVc_2 * F_v + STc_2 * F_q + PRc_2 * F_p}{SF} \tag{9}$$

where *SF* is the total weight factor and it is given as:

$$SF = F_s + F_v + F_q + F_p \tag{10}$$

Weighting factors F_s, F_v, F_q, F_p used to tune the performance of the algorithm for deciding in multi-criteria decision making algorithm are represented as real numbers with values in the boundary of [0, 1]. The length of the chromosomes is equal to four real numbers determined with accuracy to 3th decimal. In the proposed genetic algorithm only real presentation of genetic algorithms (chromosomes with real values) are used, use of binary presentation does not impose improvements in process of optimization and considering that weight factors in nature are real numbers, their conversion to binary would further complicating the whole process. The reality shows that use of GA with real numbers is usually faster than the GA with a binary presentation, because there is no need for decoding the chromosomes before the process of their evaluation by the target function. GA based on real numbers are simple and compatible combination with other optimization methods in the formation of hybrid optimization solutions keeping in mind that other optimization methods are based on real numbers [23].

The target function main goal is to modify the values derived from chromosomes in order to optimize the parameters for a given purpose. Given that future heterogeneous networks have to impose particular attention to customer satisfaction, which is mainly expressed through the Quality of Service that users receive for the proper price, a target function that directs users to networks that provide better Quality of Service has been created, taking into account all the previously defined criteria. So the first objective function is to maximize the percentage of users that are assigned to networks with higher reception signal level (US_q). The value of US_q, is taken as a simple indicator for evaluating the performance in terms of provided Quality of Service. The function that performs maximization of such a requirement is illustrated in the diagram in Figure 52.

In the process of assigning of values to weight factors, we should always keep in mind that there are certain limitations. The first limitation concerns the value of real number that must fall within strictly defined limits [0, 1]. In addition to these systemic constraints additional restrictions concerning the minimum value of each weight factor is introduced, which should be 0.1 and that should ensure that all input criteria are taken when analyzing the decision making process within the MCDM algorithm for decision making. Given these constraints, values of weight factors should be placed between the lower limit values DG = (0.1; 0.1; 0.1; 0.1) and the upper limit UG = (1, 1, 1; 1).

Creating the initial population is done by uniform random generation. Its creation is done using the function for random initial population with uniform distribution. The initial range of values of weight factors is set in the defined boundaries [0.1, 1]. As best values for the following evaluation population of 100 particles and the maximum number of 50 iterations is defined. Aim for a maximum error of all particles in the population is defined by the limit value 20.

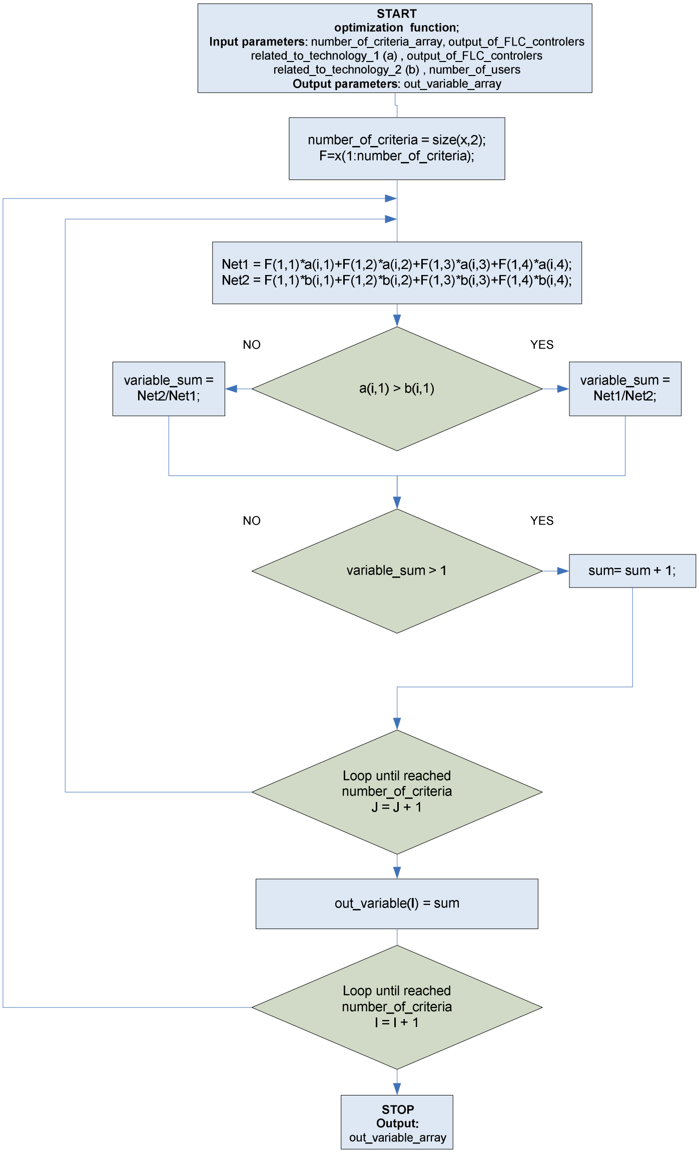

START
optimization function;
Input parameters: number_of_criteria_array, output_of_FLC_controlers related_to_technology_1 (a) , output_of_FLC_controlers related_to_technology_2 (b) , number_of_users
Output parameters: out_variable_array
number_of_criteria = size(x,2);
F=x(1:number_of_criteria);
Net1 = F(1,1)*a(i,1)+F(1,2)*a(i,2)+F(1,3)*a(i,3)+F(1,4)*a(i,4);
Net2 = F(1,1)*b(i,1)+F(1,2)*b(i,2)+F(1,3)*b(i,3)+F(1,4)*b(i,4);
NO
YES
a(i,1) > b(i,1)
variable_sum = Net2/Net1;
variable_sum = Net1/Net2;
NO
YES
variable_sum > 1
sum= sum + 1;
Loop until reached number_of_criteria J = J + 1
out_variable(I) = sum
Loop until reached number_of_criteria I = I + 1
STOP
Output:
out_variable_array

Figure 52. Diagram of the maximization function of the percentage of users that are assigned to networks with higher reception signal level

As parameters defining the process of moving the particles among space of values, the following constants are defined: cognitive acceleration c1 = 2,5 and social acceleration c2 = 2,7, acceleration of the neighborhood is not used and the corresponding parameter has a value of c3 = 1. Upon completion of the optimization optimized fuzzy controllers whose membership functions best reflect current conditions in the system of affordable technologies that build heterogeneous network are obtained.

III.3. Simulation Analysis

Taking into account the effects of different input parameters of the simulation as defined in the previous chapters, such as factor cost of access technologies from user point of view, then the factor of speed users and different number of offered services (service types defined by the appropriate bit rates and delays) simulation was set up in which all user input parameters were randomly defined, each in given value space. In addition, users receive a random simulation values for the factor cost in the range of values from 1 to 10, and a random value for the speed range in value from 1 to 10 km/h. Then, we have performed simulations for different number of mobile users, in the range from 100 to 1000 users, using increment step of 100 users. This simulation scenario represents the best approximation of realistic situation regarding radio resource management in heterogeneous networks. We have compared the results of proposed new algorithm with the well-known algorithms for radio resource management (RRM), such as: random RRM, service-based RRM and referent algorithm based on static Fuzzy Logic implementation and use of basic genetic algorithm for optimization of multi criteria decision making process, referred in this text as FGA.

The results are given in Figure 53, which gives the dependence of the user satisfaction for different number of users, when all input parameters were associated with users in random manner. The average results over the number of users are given in Figure 33. The user satisfaction means that a user is assigned to the RAT which gives optimum performance for selected user service, using the given constraints on RAT signal level, QoS, and cost for the service. The solution named M-RATS (Mobile-based Radio Access Technology Selector) uses Fuzzy controllers optimized with PSO (Particle Swarm Optimization) and Genetic Algorithm for optimization of parameters of MCDM as described in the previous sections. Simulation results are given for different number of users in scenario, where users are randomly distributed within the given services area, which consists of WWAN and WLAN cells. Even with less users in the scenario (e.g., 100 or 200 users, as it can be seen from Figure 53), the M-RATS algorithm shows better behavior than well-known RRM techniques, including referent FGA, random-based RRM, and service-based RRM.

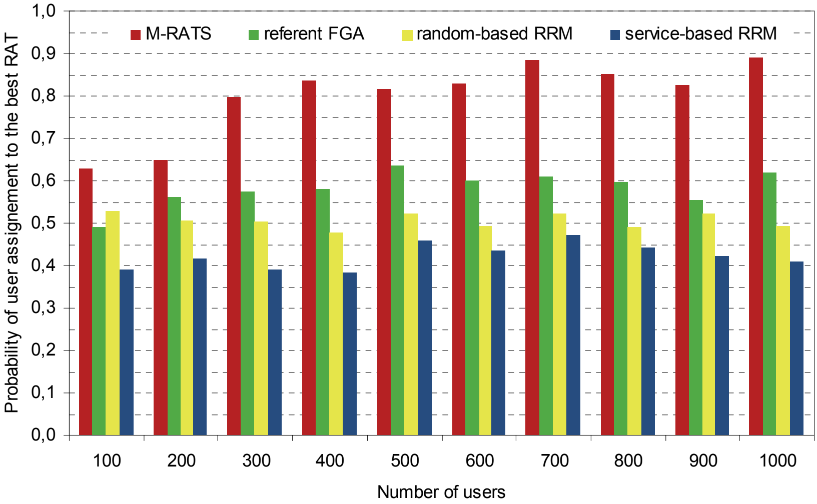


Figure 53. Probability for user access via preferred RAT (user satisfaction), using uniform random distribution of the input parameters to the algorithm

The results regarding the probability of assignment of users to the best RAT (which is subjective for each user) goes between 80 and 90% for all simulations with more than 300 users, because there are more users participating in the process.

Most of the systems for selection of network access are based on one criterion. In contrast to such approaches, the proposed mechanism M-RATS is a multicriteria algorithm that seeks to satisfy different requirements or objectives set by various criteria.

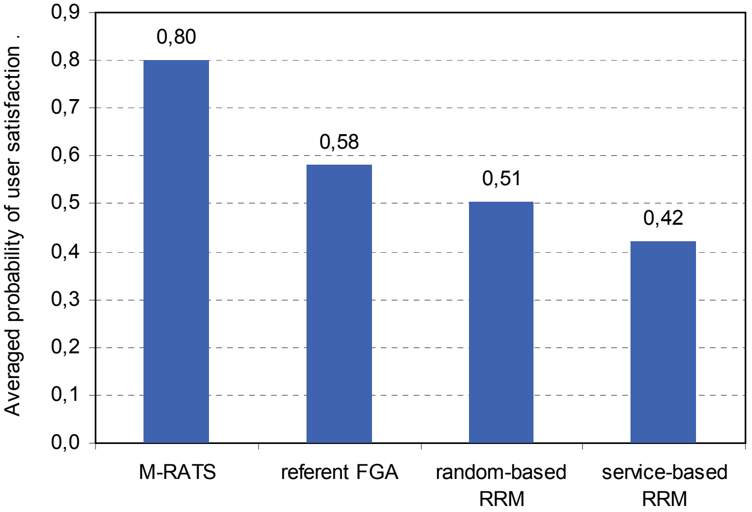


Figure 54. Average probabability for user satisfaction averaged over number of users

All previous algorithms do not address the roles of different sides in the selection process (users and network providers) and do not provide a complete solution that can be applied in heterogeneous networks. This is particularly important because of the nature of the network selection process in which both, the operators and the user, want to control the network selection process and their roles in it should be strictly defined and in accordance with the degree of interdependence between different radio access technologies that consist the architecture of a heterogeneous network. Our approach allows defining the different roles of all parties in the decision-making in general terms that increases their satisfaction by the proposed final choice.

In heterogeneous wireless architectures consisted of wireless networks which are owned by different entities (i.e., network operators), the user should be able to control the RAT selection, assisted by network entities n the service stratum, which may belong to a third party. In loosely-coupled or tight-coupled architectures of heterogeneous networks owned by single operator, the RAT selection model can be implemented as a system which is divided in two modules, one set in the user terminal and the other set on the network side as an integral part of the mechanism for joint radio resource management.

The goal of the proposed mechanism, which is in fact the final choice of radio access technology, is done by direct interaction between the two modules.

Conclusion

In this chapter a newly designed method for selection of radio access technology, based on algorithms from artificial intelligence was designed and presented. The new algorithm is based on mechanisms for optimization using particle swarm, genetic algorithms, fuzzy logic and multi-criteria decision making methods in order to obtain an optimal system for intelligent decision when choosing the wireless access network.

The system is applicable in environments which require RAT decisions based on diverse parameters and criteria using previously stored history obtained from the usage of a given service in a given network. Such data can be stored either at the mobile terminal or in centralized network node. The proposed new algorithm is applied as a part of the algorithm for initial selection of the access network as well as for vertical handover control. Using the proposed system, based on continuous monitoring of services and key performance indicators, algorithm provides network selection that best meets the common quality requirements determined by measured indicators and network parameters, including the requirements from the user preferences and those set by the network operators. The input criteria are speed of the mobile terminal, Quality of Service, type of service and cost. The approach can be easily extended to additional criteria when needed. Simulation analysis and comparison of the proposed algorithm with well-known mechanisms in heterogeneous wireless environment have shown that it outperforms other mechanisms, providing highest probability for mobile user satisfaction.

However, the proposed mobile-based radio access technology selection algorithm requires higher computational power from mobile terminals. On the other side, development of the mobile terminals paves the ground that next generation wireless networks will be user-centric, with multimode mobile devices with capabilities to implement the given intelligent RAT selection algorithm.

Chapter IV

M-RATS: MOBILE-BASED RADIO ACCESS TECHNOLOGY SELECTOR FOR HETEROGENEOUS WIRELESS ENVIRONMENT

Wireless Network Selection Algorithm

In this chapter we provide final description of an innovative novel algorithm for radio networks selection in heterogeneous environment, which is created using biologically inspired algorithms. Considering all building blocks, and algorithms, presented in previous chapters in the following part detailed description of the final algorithm is presented. The algorithm consists of four building components as shown in Figure 1. The first component or module is a set of parallel Fuzzy Logic (FL) controllers, which as input has the measurements data for different selection criteria, including user requirements, QoS requirements, service policies, as well as radio link conditions in different wireless technologies present in the user's area. The second module is multi-criteria decision mechanism algorithm, which uses as inputs the outputs of the FL controllers form the first module. The third module is Genetic Algorithm, which does optimization of weighting coefficients of different input criteria. That is, each criterion can have different weight, which depends upon the assumption of its impact on the best network selection process (i.e. the decision). The fourth module is Particle Swarm Optimization (PSO) [16], [17], mechanism which dynamically modifies the functions of FL controllers in the first module (shown in Figure 55).

This system is not limited only to access network selection, but it can be also used for solving other optimization problems as well. The proposed scheme in this book is targeted for wireless networks selection in heterogeneous environment, so the decision as an outcome should select the best wireless network (among all present at the moment for a given user) or to rank in certain order all present radio access networks in certain order.

Initial phase of this scheme is data collection, by using measurement of the parameters of the Radio Access Networks (RAN). This process includes operator and user preferences. For instance, signal power and signal to noise ration on link layer are received as measured parameters from each network.

Then, additional preferences can be service cost over a given RAN, which depends upon Service Level Agreements (SLA) between the user and each of the wireless networks i.e. depends upon service policies.

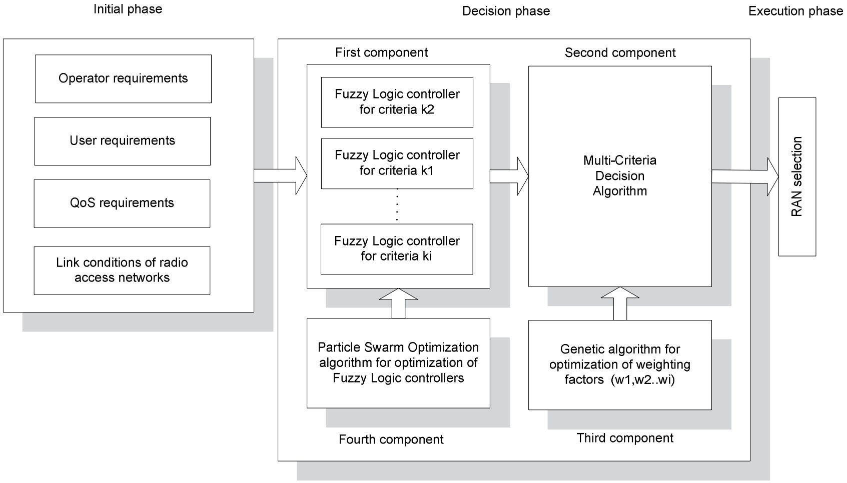

Initial phase
Decision phase
Execution phase
Operator requirements
User requirements
QoS requirements
Link conditions of radio access networks
First component
Fuzzy Logic controller for criteria k2
Fuzzy Logic controller for criteria k1
Fuzzy Logic controller for criteria ki
Second component
Multi-Criteria Decision Algorithm
RAN selection
Particle Swarm Optimization algorithm for optimization of Fuzzy Logic controllers
Genetic algorithm for optimization of weighting factors (w1,w2..wi)
Fourth component
Third component

Figure 55. Radio Access Network selection scheme

The proposed scheme assumes that mobile terminal has enough processing power, memory capacity and battery support, so it can provide functionalities described above. By the year 2020 mobile phones will probably have processing capabilities of current power computers, so, they can operate a database, do processing on given time intervals using measurements data from a given timeframe in the past.

IV.1. Initial History Buffer Size

Crucial part of selection mechanism in this terminal-controlled scenario is providing adequate data as an initial training sequence of Artificial Intelligence (AI) algorithms in the solving scheme. There are several possibilities to achieve this goal. Data can be provided by the network in a loosely-coupled scenario considering operator assisted terminal-controlled scenario, or by imposing training period in terminal itself considering pure terminal-controlled scenario. Never the less quantity of historical data that have to be included in the initial phase of training of AI algorithms must be determined.

In order to define the initial buffer size several simulations have been conducted with different size of initial buffer and comparative analysis has been made. As it can be seen on Figure 56, best buffer size can be defined in borders of 300- 400 historical intervals.

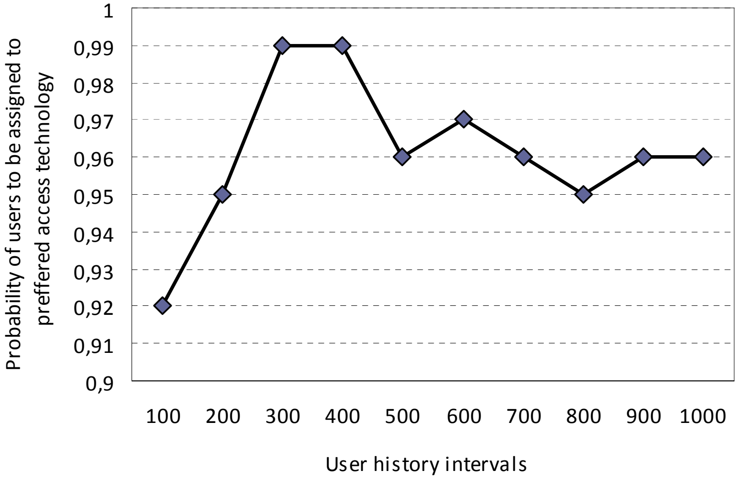


Figure 56. Initial buffer size dependency

History intervals as shown in Figure 56, are time periods between subsequent measurements of the RATs parameters of the mobile terminals as well as the snapshot of user and service network demands in that period of time. They present most recent history of the user activity as well as user personal and user service demands from the network.

IV.2. Simulation analysis

First necessary step to create simulation environment is to identify several models that defines the tested reality in the simulation and these are: the system model, the model for the mobility of users, propagation model, traffic model and service model. The first three models are part of the software package [24], while the other two models are defined in accordance with the traffic theory and according to the type of services that are most common or suspected to be most common in future heterogeneous networks. The following Figure 36, shows the simulation system that covers the used system and service models in the evaluation of the proposed algorithm.

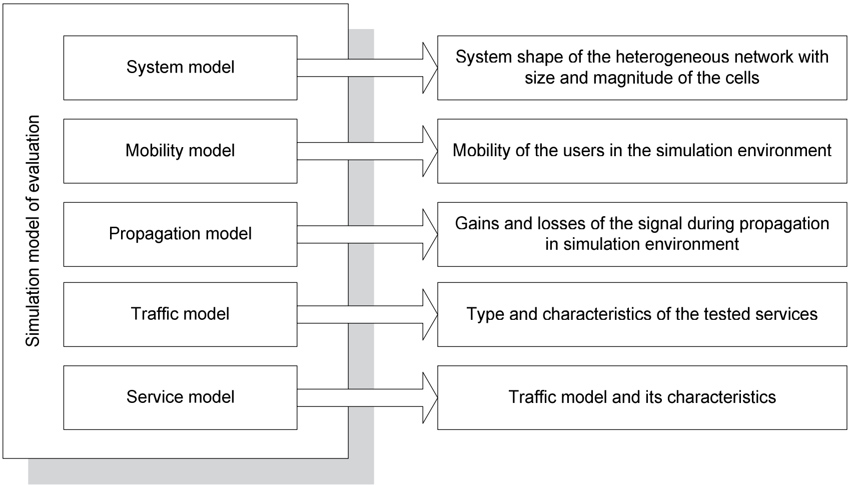

Simulation model of evaluation
System model
Mobility model
Propagation model
Traffic model
Service model
System shape of the heterogeneous network with size and magnitude of the cells
Mobility of the users in the simulation environment
Gains and losses of the signal during propagation in simulation environment
Type and characteristics of the tested services
Traffic model and its characteristics

Figure 57. Simulation models in the tested environment

For the purpose of simulation analysis we have used the following mobility model, which provides randomness for user mobility. Each mobile terminal has velocity calculated according to the following:

$$v_i = v_{i-1} * C_v + \sqrt{1 - C_v^{\,2}} * v_{mean} * N \tag{11}$$

where v$_i$ is the user speed [m/s]. C_v is the correlation of the velocity between time steps. It depends on both a_{mean} that is the mean acceleration of the mobile user and v_{mean} which is mean velocity of mobile user. C$_v$ is calculated as:

$$C_v = \left(\frac{-dt * a_{mean}}{v_{mean}} \right) \tag{12}$$

where N is Rayleigh distributed magnitude with mean 1 and a random direction. v_{mean} is the mean speed of mobiles. v_{mean} was set to 10 km/h and a_{mean}

has been set to 1 km/h^2, which are typical values for urban environment. Figure 58, shows the users, in marked spots, over the simulated environment.

The service model represents the third aspect modeled by the proposed simulation method. He is defining the type of services and also defines their representation among the users in the system. In the proposed form service model predicts the existence of four services that are defined by its required bit rate and time of propagation. The first service type is defined by a low bit rate and small propagation time (Round Trip Time - RTT) and is used for handling of voice services, second service type is defined by medium bit rate and low propagation time and is used for services such as video telephony, the third service type is defined by the average bit rate and average time of propagation and is used for services such as video streaming, fourth service type is defined with a high bit rate and can handle bigger time propagation and is used for data services that do not strictly defined time frames.

In the following part of this chapter we show the modeling of the four types of services used in the simulation analysis. Each of the four service types is described and defined by requested bandwidth (in bps) and latency (in ms).

First service type is defined with lower bandwidth (data rate) and lower latency, because it includes telephony services. The second service type is defined with the higher average data rate and relatively small latency, and it is targeted to video conferencing service.

The third service type is video streaming, which requires higher bandwidth per connection, but can accommodate higher latencies compared to telephony. The last one, fourth service type, models non-real time services, such as web and email, and it is defined with fairly higher data rates (since these services use TCP, which consumes all available bandwidth end to end). All four service types, for the purpose of simulation analysis, are defined with the following pairs of values (bandwidth, latency), respectively:

[service_latency (ms), service_bandwith (kbps)]

$$\{[100, 64]; [200,128]; [400,256]; [800,512]\} \tag{13}$$

During the simulation for a given number of active users N, each user is randomly assigned to one of the four types of services defined above. Traffic model reflects the process of creating and duration of user services defined by service model. Its goal is to give a realistic picture of the user traffic within the simulation. For each service, the connection duration is modeled with the Poisson process. The mean holding time is set to 50 seconds, based on comparative measurements.

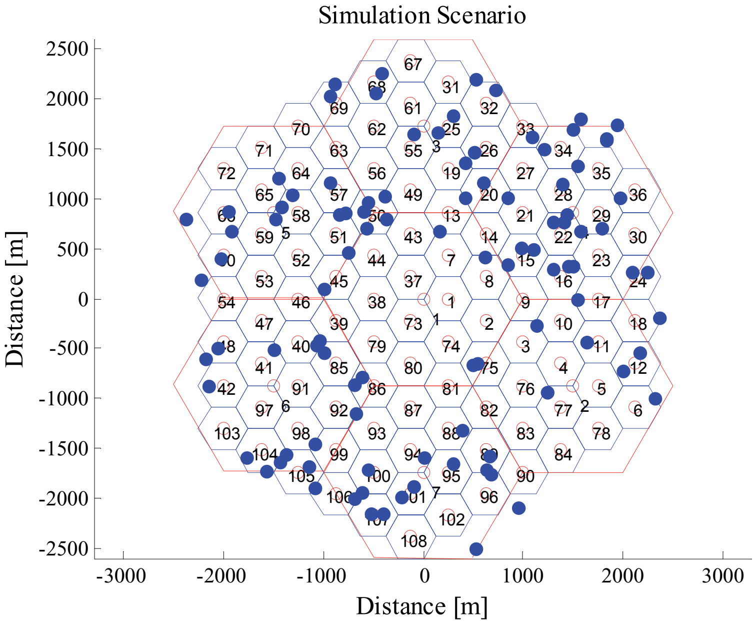


Figure 58. Simulation Scenario

The performance of wireless communication system largely depends on the situation in the mobile radio channel. Propagation of radio waves through the wireless environment involves the implementation of simulation techniques for various natural phenomena such as reflection, diffraction and scattering. Model of propagation simulates various gains and losses during propagation of radio waves between the transmitter and receiver within the defined Simulation environment. Propagation model used in simulation in logarithmic form is defined by the following formula:

$$G = G_D + G_F + G_R + G_A \qquad (14)$$

As it can be seen from the formula, gain G depends from following factors: G_D that represents attenuation due to wave propagation through the air environment, G_F that represents attenuation due to shadow fading, G_A that represents the gain of the system introduced by receiving and transmitting antenna and G_R that represents a loss due to a Rayleigh fading. In Figure 59, propagation of signal from neighboring BTS's as a result from propagation model is presented.

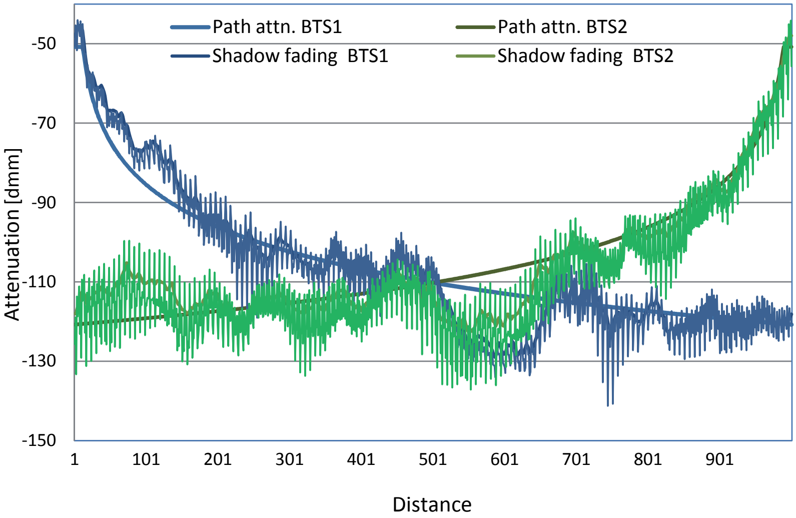


Figure 59. Propagation of signal from 2 BTS's as a result from defined propagation model

The first step in the simulation starts with defining of fuzzy logic controllers to fit the FL controller parallel scheme given in Fig.1. In the simulation scenario provided, two access technologies (WWAN and WLAN), and all appropriate algorithms are designed for the representation of heterogeneous network consisting of the proposed technologies. Outputs from fuzzy logic controllers represent the degree of membership to each of the entries in terms of scaled fuzzy logic rules. For optimization of fuzzy logic controllers, algorithms based on particle swarm are used. Their ultimate goal is to generate a set of optimized fuzzy logic controllers that best reflect the required goals, in order to optimize FLC, where membership function are tuned to the measured signal strengths and wanted user behavior. Considering that two RAT technologies (3G network and WLAN) are analyzed in the scenario we have two outputs from each FL controller.

In our simulations the PSO algorithm uses swarm size of 50 particles while maximum number of iteration is set to 50. Evaluation function is based on minimizing the mean square error (MSE) while comparing it to the expected predefined values. Expected values are defined as values taken from humanly decision that would be made if access network selection is done by human for every point in time and separately for each analyzed criteria [16].

Considering the evaluation done in section III, for Initial data buffer in the simulations we use range of 300 historical intervals. This means that in the scenario training period of 300 time slots (intervals) is imposed and initial data gather during this period are stored. These data are then used in GA for acquiring the first set of weights that will be imposed in the decision done using the multi-criteria approach. After this initial moment, in every other step, data from the buffer are constantly refreshed with new measured data. This is done in FIFO manner, first input data from training period are replaced and pushed out from the buffer, all other data are moved one place backwards and on the last place newly measured data are placed. During this process buffer data size remains the same 300 historical measured data, so we have a certain time window for data collection in the mobile terminal, obtained from different present RATs (in this case, there are two RATs, i.e., 3G and WLAN).

Genetic Algorithm (GA) is used as optimization method for determination of appropriate values for weights of different criteria in multi-criteria decision making approach. The goal of the GA is to optimize the weights upon locations of the users and their demands to the network (which is dependent upon service type initiated by each user). With this approach, the GA can assign weighting coefficient to provide best user satisfaction. In our analysis we use 200 iterations, which is based on the fact that there is no improvement after successive 100 generations in most cases.

On the following Figure 60, results from the goal function during the process of optimization of MCDM weighting factors are presented.

In order to address the process of initial selection of an access network a new evaluation mechanism that is independent of other systems for radio resource management was set. Mechanism determines the percentage of users who are assigned to networks with better quality parameters of the service.

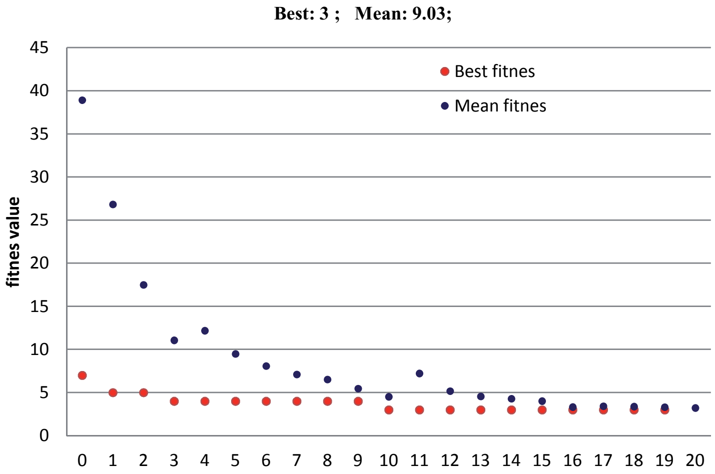


Figure 60. Values of the goal function during the optimization process

This mechanism addresses the performance of selection in the selection process for access to the heterogeneous network by users in terms of the achieved quality of service. This implies that the chosen network has better quality characteristics (level of reception signal, bit rate and delay from end to end) for a given service (voice, video calling, video streaming or data traffic) to the user at a given location and speed of movement. If as a result of the initial user selection is assigned to the network with better quality characteristics, then we say that the user is satisfied, and the ratio of satisfied users in relation to the total number of users in the simulation scenario is called coefficient of satisfied customers in terms of total number of users and is denoted by (Qp).

In Figure 61, we have compared the proposed algorithm with almost all other relevant algorithms by means of evaluation criteria Qp. The worst results are obtained with mobility based Radio Resource Management (RRM), while the best results are obtained with our proposed algorithm which includes FL optimized with PSO and GA for Multi-Criteria Decision Making (MCDM).

FGA stands for Fuzzy-logic Genetic Algorithm, while Sigmoidal refers to membership functions for FL controllers, which are trapezoidal or triangular in other cases. Several runs of simulation have been carried out for different number of users in simulation scenario (from 100 to 1000 users), and the results show the probability that a user is attached to RAT which provides the best user satisfaction regarding the bandwidth and latency requirement for each service type.

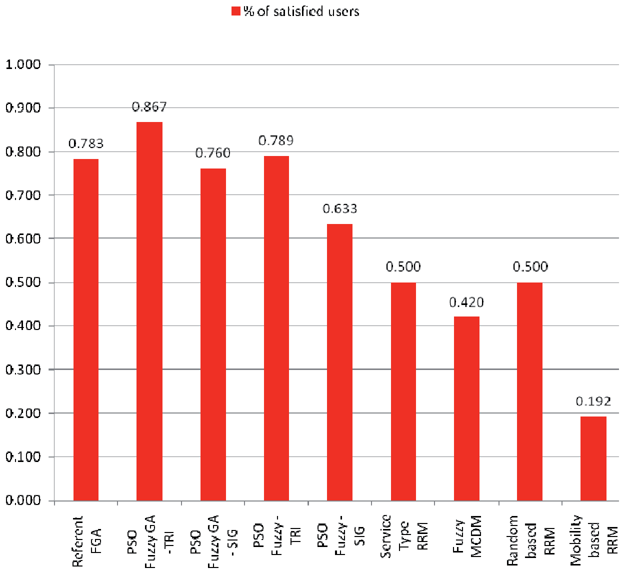


Figure 61. Arithmetical mean value of coefficients that represents the user satisfaction from selection process

The proposed wireless access selection algorithm in this paper achieves around 40% enhancement compared to service-based and random-based selection algorithms. It also shows better results from all other wireless network selection algorithms, which is more noticeable when the number of data samples (users) is higher.

To give a complete review of performance and features of the proposed algorithm besides the above modifications to the building blocks of the same elements and the changes of parameters in the simulation were analyzed. Namely, in the presented simulation, new results that include analysis based on different number of users in the scenario as well as changes of various building elements of the algorithm for selecting the access network, such as changing the parameters of fuzzy logic controllers and inclusion or exclusion of certain building elements of the proposed algorithm were conducted. In this way basic feature of the new algorithm performances were determined.

Besides the analyzed aspects of the simulation there are additional elements that affect its performance, having in mind that basic input parameters of the proposed algorithm include: the level of signal reception of radio access technologies, user requirements reflected by the demand factor cost, operator requirements to guide users to a given technology depending on the speed of the terminal, the qualitative requirements of the services offered by heterogeneous networks.

The purpose of the performance analysis of the proposed algorithm is to perform the analysis of the dependency of the algorithm from different input variables. This involves setting the Simulation environment in which as the first would be changing the speed of movement of the terminal in the range covered by the input fuzzy logic controller (1 km/h to 10 km/h). This simulation scenario is presented on Figure 62.

The second change is in the number of services offered to mobile subscribers in this scenario from 1 service to the maximum number of four different services presented on Figure 63.

Third shift factor is factor of cost related to access network determined by the user in the range of 1 to 10 units that would reflect the willingness of consumers to pay for services received while setting a different number of mobile users in the simulation as presented on Figure 64.

Considering that the best performance by presenting modifications to the proposed algorithm gives M - RATS algorithm (which presents an algorithm based on PSO fuzzy logic optimized controllers with triangular shape and MCDM algorithm for decision making optimized using a genetic algorithm) in further comparisons to determine the suitability of structured algorithm and to evaluate its performance, as basic algorithms for comparing him to take classical radio resource algorithms based on random selection and a defined service type, and a modified version of the intelligent algorithm with a basic minimum set of elements using the same fixed input fuzzy logical controllers with triangular shape and MCDM algorithm optimized by genetic algorithm. Such a comparison would provide a realistic image of actual performance of the proposed algorithm. Results from simulations for changing of the parameters are presented in the following figures.

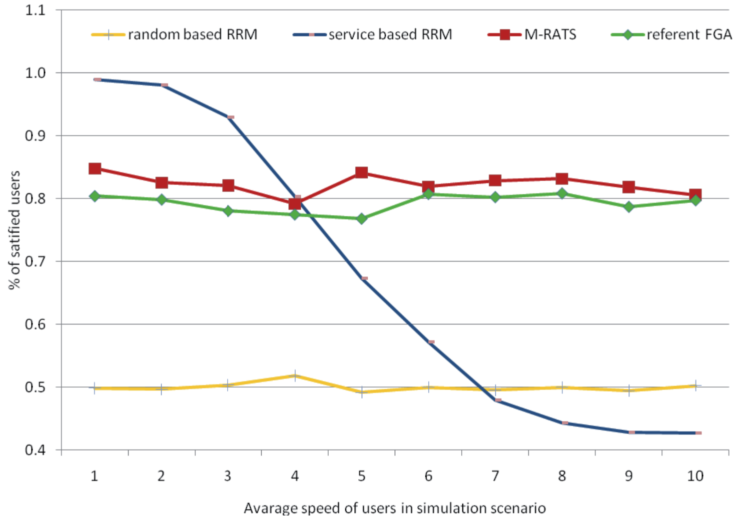


Figure 62. Dependency of the algorithm regarding parameter of user speed

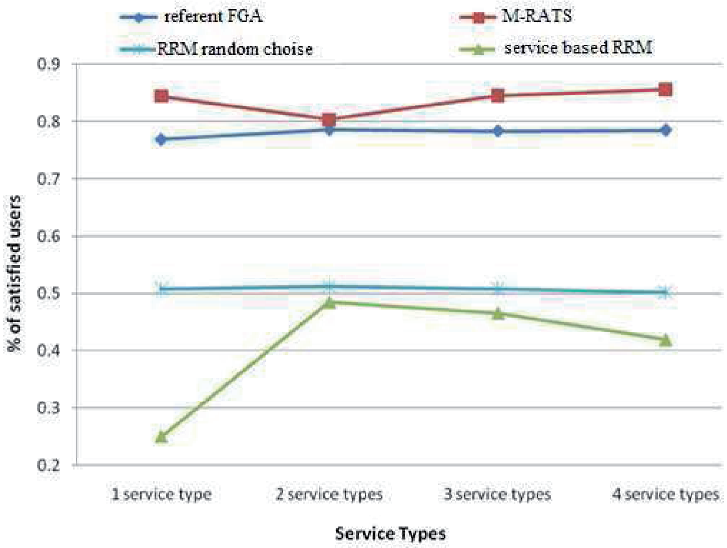


Figure 63. Dependency of the algorithm regarding number of service types given to the users in scenario

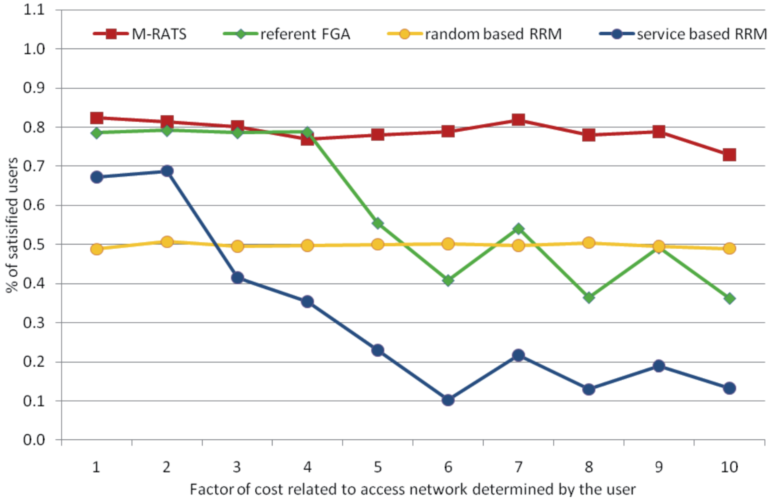


Figure 64. Dependency of the algorithm regarding factor of user cost

It could be noted that in some segments, for certain simulation parameters, specific algorithms for selecting the access network have better performance than the proposed algorithm, due to the uniform distribution of input parameters with fixed values of all input variables except those whose impact is analyzed. This is illustrated with an analysis regarding the impact of speed of movement of mobile terminals based algorithm for selecting the access network, where for fixed variables mechanism for radio resource management based on service type had better features for more static users at factor cost of the fixed value of 3. Generally we can say that each input variable has its effect on the mechanism for selecting an access network by themselves are not as significant as when they are viewed in the context of the values of other parameters. It is this advantage of the algorithms based on artificial intelligence that such interdependence is better reflected compared to deterministic algorithms. In real scenario where all variables are random proposed intelligent algorithm give better results, and confirmation of this conclusion is given by the simulation described in this section.

Chapter V

CONCLUSIONS

In accordance with the latest research, this book presents describes the direction and techniques together with actual implementation of new methods for controlling and improving of the quality of services in future generation mobile networks. It gives overview of the architectures of the future generations of networks, it performs analysis of network architectures in heterogeneous networks and the ways radio resource management is achieved in each of them. Main focus was given on the heterogeneous wireless networks, and a new proposal is presented for their architecture of internetwork operation using the Internet model for interoperability between radio access technologies that participate in the construction of heterogeneous networks. In this proposed architecture all relevant aspects for its proper functioning are analyzed. Analyses were conducted on the design of network-level transport of packets through the system of radio technologies, the procedures for authentication of users and the procedures for authorization and users and transmission technologies. Analyses were conducted on the mechanisms for continuous monitoring of quality of customer service that is based on a method for routing policies of the application package within the architecture.

The proposed new architecture in this text defines the unique ecosystem of inter-network operation between different RAT’s that form heterogeneous wireless networks. Main goal of proposed new solution is to raise the quality of services offered along with user preferences and network in the center of the network architecture. This implies that all basic functional systems of the architecture, routing of the packets per application or service, handover, initial choice of network access and so on, work in close relation with the control module of user services. In this context, an analysis of key performance indicators for consumer services that accurately define the quality of services provided to a user of packet based (IP) services in mobile/wireless environment has been performed. Measurement architectures were reviewed to measure basic performance indicators and proposed ways of their continuous measurement. In this context, foundations were laid for a new dynamic protocol for exchanging information polices (QoSPRo) between elements of architecture in the form of protocol procedures.

As part of the proposed architecture processes of authentication and authorization when connecting to the access technologies were analyzed and innovative techniques were introduced. In this context as the basis of defined authentication process an independent mechanism for authentication (MIM) for authentication procedures based on certificates was set. The process of authentication and authorization is based on PANA protocol as the vanguard of establishing authenticated IPSec tunnel, which is fundament in the design of the proposed architecture for interworking in heterogeneous networks.

Furthermore, the book defines a new algorithm for selecting of access network in heterogeneous wireless environment by using algorithms from artificial intelligence. In this direction at first, the impact on network architecture leaving some building blocks of radio resource management are covered and defined the role and impact of the method for selecting an access network on the overall quality of the service of the user and his sense of network quality. A new method for selecting an access network based on artificial intelligence is proposed as a very important segment of the overall radio resource management in future heterogeneous networks. The proposed algorithm is composed of: an algorithm for optimization using particle swarm, genetic algorithms, fuzzy logic and Multi Criteria Decision Making methods in order to obtain an optimal system for intelligent decision in process of access network selection. The presented algorithm, called "M-RATS", was placed in Simulation environment and an analysis of its performance in terms of customer satisfaction was made.

In simulation, effects from heterogeneous environment were identified for each of the individual elements of the system on the final results as compared to the reference algorithms. The results showed that the proposed algorithm gives superior results in terms of comparative algorithms and as such is the first choice for implementation of a system for selecting an access network in future generations of wireless heterogeneous networks. The functionality of the algorithm and its construction has made quite flexible and usable procedures for network selection and the initial choice of access technology. The system is applicable in environments where there is a need to adopt a decision based on diverse parameters and criteria for making decisions based on their previous history. The proposed new algorithm presents suitable application as part of the initial algorithm for selecting of access network as part of the control system handover. This system, thanks to the continuous monitoring of services and information, for key performance indicators, proposes to choose network technology that would best meet the common quality requirements under certain criteria: the measured parameters of network performance indicators, requests by the user and preferences by the network operator. In this way, a complete image of future generation of network technologies is given, that besides meeting the basic concept of heterogeneity need to maintain the satisfactory level of customer quality of service is given.

The proposed new solutions and innovations give an important contribution to future generations of mobile networks, which will be fully IP based. In such environments customer service will be separated from transport technologies and customer satisfaction of various services in heterogeneous wireless and mobile environment will be their primary goal.

REFERENCES

[1] T. Janevski, “Traffic Analysis and Design of Wireless IP Networks”, Artech House Inc., Boston, USA, 2003.

[2] ITU-T, Y.2001, “General overview of NGN”, December 2004.

[3] ITU-T, Y-2002, “Overview of ubiquitous networking and of its support in NGN”, October 2009.

[4] Toni Janevski, “5G Mobile Phone Concept”, IEEE Consumer Communications and Networking Conference (CCNC) 2009, Las Vegas, USA, January 2009.

[5] M. Kassar, B. Kervella, G. Pujolle, ”An overview of vertical handover decision strategies in heterogeneous wireless networks”, Elsevier Computer Communications 31, p.2607-2620, 2008.

[6] W. Luo, E. Bodanese, “Optimizing Radio Access in a Heterogeneous Wireless Network Environment”, IEEE International Conference on Communications, Dresden, Germany, 14-18 June 2009.

[7] M. Ha Nguyen Tran Hasegawa, Y. Murata, H. Harada, “Representation of user satisfaction and fairness evaluation for user-centric dynamic spectrum access”, Personal, Indoor and Mobile Radio Communications (PIMRC), Tokyo, Japan, 13-16 September 2009.

[8] J. Perez-Romero, O. Sallent, R. Agusti, “A Novel Metric for Context-Aware RAT Selection in Wireless Multi-Access Systems”, ICC’07, Glasgow, Scotland, 24-28 June 2007.

[9] A. Tudzarov, T. Janevski, “M-RATS: Mobile-based Radio Access Technology Selector for Heterogeneous Wireless Environment“, Telfor 2010, Belgrade, Serbia, November 23-25, 2010.

[10] ITU-T, Y.2173, “Management of performance measurement for NGN”, September 2008.

[11] T. Janevski, A. Tudzarov, M. Porjazoski, P. Latkoski, “System for Analyses of End-to-End Quality of Data Services in Cellular Networks”, IEEE Eurocon 2009, Saint Petersburg, Russia, May 18-23, 2009.

[12] T. Janevski, A. Tudzarov, I. Efnushev, P. Latkoski, M. Porjazoski, D. Gjorgjiev, “Applicative Quality of Testing System Software for IP-based

Services in Mobile Networks”, IEEE SoftCOM 2007, Split-Dubrovnik, Croatia, September 27-29, 2007.

[13] M.Alkhawlani and A.Ayesh, “Access Network Selection Based on Fuzzy Logic and Genetic Algorithms”, Advances in Artificial Intelligence, Volume 8, Issue 1, January 2008.

[14] J.Zander and S. Kim, “Radio Resource Management for Wireless Networks”, Artech House, Boston, Mass, USA, 2001.

[15] J. Kennedy, R.C. Eberhart, “Particle swarm optimization”, Proceedings of the IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948, 1995.

[16] M. Clerc, J. Kennedy, “The particle swarm explosion, stability, and convergence in a multidimensional complex space”, IEEE Transaction on Evolutionary Computation, vol.6, pp. 58–73, 2002.

[17] J. Perez-Romero, O. Sallent, R. Agustı, and M. A. Dıaz-Guerra, “Radio Resource Management Strategies in UMTS”, John Wiley & Sons, New York, NY, USA, 2005.

[18] L. Giupponi, R. Agustı, J. Perez Romero, and O. Sallent, “A novel joint radio resource management approach with reinforcement learning mechanisms,” in Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference (IPCCC ’05), pp. 621–626, Phoenix, Arizona, USA, April 2005.

[19] R. Agustı, O. Sallent, J. Perez-Romero, and L. Giupponi, “A fuzzy neural based approach for joint radio resource management in a B3G framework,” in Proceedings of the 1st International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE ’04), pp. 216–224, Dallas, Tex, USA, October 2004.

[20] A.Wilson, A. Lenaghan, and R. Malyan, “Optimizing wireless access network selection to maintain QoS in heterogeneous wireless environments,” in Proceedings of the Wireless Personal Multimedia Communications (WPMC ’05), Aalborg, Denmark, September 2005.

[21] J. Martin, A. Eltawil. “Towards a Unified Wireless Network Involving Reconfigurable Devices”, by Clemson University, May, 2010.

[22] M. Sugeno, “Industrial Applications of Fuzzy Control”, Elsevier Science Ltd, 1985.

[23] R. Haupt and S. Haupt, “Practical Genetic Algorithms”, John Wiley and Sons, 2nd Edition, 2004.

[24] MathWorks, Inc., Genetic Algorithm and Direct Search Toolbox Users Guide, the MathWorks, Inc, Version 2.0.1, Matlab Release 2006a, March 2006

[25] Poonam Arora, Prem R. Vemuganti, Praveen Allani, “Comparison of VPN Protocols – IPSec, PPTP, and L2TP”, Department of Electrical and Computer Engineering George Mason University, Project Report ECE 646 (Fall 2001)

[26] Hamzeh, K. et al, “Point-to-Point Tunneling Protocol (PPTP)”, RFC 2637, July 1999.

[27] Townsley, W. et al, “Layer Two Tunneling Protocol - L2TP”, RFC 2661, August 1999.

[28] Cisco Systems, Inc. and its subsidiaries (including Cisco Consumer Products LLC, Cisco WebEx LLC, and Pure Digital Technologies LLC) (collectively “Cisco”)

[29] Microsoft Corporation (Redmond, WA).

[30] Kent, S. and Atkinson, R., “Security Architecture for the Internet Protocol”, RFC 2401, November 1998.

[31] Oppliger, R., “Security Technologies for the World Wide Web”, Artech House Computer Library, 2000.

[32] Yuan, R. and Strayer, T., “Virtual Private Networks”, Addison-Wesley, 2001.

[33] D. Forsberg, Y. Ohba, B. Patil, H. Tschofenig, A. Yegin, “Protocol for Carrying Authentication for Network Access (PANA)”, draft-ietf-panapana- 07 (work in progress), December 2004.

[34] A Olivereau, A.F. Gomez Skaremta, R.M Lopez, B Weyl, P. Brandao, P Mishra, c. Hauser, “An advanced Authorization Framework for IP-based B3G Systems”, 14th IST Mobile & Wireless Communication Summit, Dresden 19-23 June 2005

[35] Toni Janevski, Aleksandar Tudzarov, Marko Porjazoski, Pero Latkoski, University “Sv. Kiril i Metodij”, Faculty of Electrical Engineering and

Information Technologies, “System for analyses of end-to-end quality of data services in cellular networks”,

[36] Toni Janevski, Aleksandar Tudzarov, Marko Porjazoski, Pero Latkoski, Ilija Efnushev, Gjorgi Madzarov, Dejan Gjorgiev, “Design of Applicative Quality Testing System for Data Services in Mobile Networks”, Wireless World Congress (WWC), San Francisco, USA, May 14-16, 2008.

[37] Toni Janevski, Aleksandar Tudzarov, Marko Porjazoski, Pero Latkoski, Ilija Efnushev, Gjorgi Madzarov, Dejan Gjorgiev, University “Sv. Kiril i Metodij”, Faculty of Electrical Engineering and Information Technologies, “Applicative Solution for Testing the Quality of Data Services in Mobile Networks”, IEEE Melecon 2008, Ajacio, France, May 5-7, 2008.

[38] Toni Janevski, Aleksandar Tudzarov, Marko Porjazoski, Pero Latkoski, Ilija Efnushev, Dejan Gjorgiev,, “Applicative Quality of Testing System Software for IP-based Services in Mobile Networks”, IEEE SoftCOM 2007, Split-Dubrovnik, Croatia, September 27-29, 2007.

[39] Recommendation ITU-T Y.2173, SERIES Y: Global information infrastructure, internet protocol aspects and next-generation networks; Next Generation Networks – Quality of Service and performance; Management of performance measurement for NGN.

[40] 3GPP Technical Specifications TR 25.881, “Improvement of RRM across RNS and RNS/BSS (Release 5), v5.0.0, Dec. 2001

[41] Toni Janevski, Aleksandar Tudzarov, Dusko Temkov, “Modeling of TCP and UDP Internet Traffic with Middle-Level Self Similarity”, in the Proceedings of ICEST 2004, Bitola, Macedonia, June 16-19, 2004.

[42] Aleksandar Tudzarov, Dusko Temkov, Toni Janevski, Ognen Firfov, “Empirical Modeling of Internet Traffic at Middle-level Burstiness”, IEEE Melecon 2004, pp.535-538, Dubrovnik, Croatia, May 12-15, 2004.

[43] Toni Janevski, Dusko Temkov, Aleksandar Tudzarov, “Statistical Analysis and Modeling of the Internet Traffic”, ICEST 2003, pp.170-173, Sofia, Bulgaria, October 16-18, 2003.

[44] Aleksandar Tudzarov, Dusko Temkov, Toni Janevski, “Internet Traffic Statistics and Models”, pp.T-60-T-65, ETAI VI - National Conference with International Participation, Ohrid, Macedonia, September 17-20, 2003.

OTHER PUBLICATIONS ON THE SAME TOPIC

Papers published in journals

- **Aleksandar Tudzarov**, Toni Janevski, “Design for 5G Mobile Network Architecture”, International Journal of Communication Networks and Information Security, to appear June 2011.
- **Aleksandar Tudzarov**, Toni Janevski, “Efficient Radio Access Technology Selection for the Next Generation Wireless Networks”, International Journal of Research and Reviews in Next Generation Networks, April 2011.
- **Aleksandar Tudzarov**, Toni Janevski, “Automatic Wireless Network Selection by using Naturally Expired Algorithms”, Int’l Transactions on Computer Science and Engineering, Vol.57, No.1, December 2009.

Papers at conferences

- **Aleksandar Tudzarov**, Toni Janevski, “Experience-based Radio Access Technology Selection in Wireless Environment”, IEEE Eurocon 2011, Lisbon, Portugal, 27-29 April 2011.
- **Aleksandar Tudzarov**, Toni Janevski, “M-RATS: Mobile-based Radio Access Technology Selector for Heterogeneous Wireless Environment“, Telfor 2010, Belgrade, Serbia, November 23-25, 2010.
- Toni Janevski, **Aleksandar Tudzarov**, Marko Porjazoski, Pero Latkoski, “System for Analyses of End-to-End Quality of Data Services in Cellular Networks”, IEEE Eurocon 2009, Saint Petersburg, Russia, May 18-23, 2009.
- Toni Janevski, **Aleksandar Tudzarov**, Pero Latkoski, Marko Porjazoski, Ilija Efnushev, Gjorgji Madzarov, Dejan Gjorgjiev, “Design of Applicative Quality Testing System for Data Services in Mobile Networks”, Global Mobile Congress 2009, Shanghai, China, October 12-14, 2009.
- Toni Janevski, **Aleksandar Tudzarov**, Pero Latkoski, Marko Porjazoski, Ilija Efnushev, Gjorgji Madzarov, Dejan Gjorgjiev, “Applicative Solution for Testing the Quality of Data Services in Mobile Networks”, IEEE Melecon 2008, Ajacio, France, May 5-7, 2008.

- Toni Janevski, **Aleksandar Tudzarov**, Ilija Efnushev, Pero Latkoski, Marko Porjazoski, Дејан Ѓорѓиев, “Applicative Quality of Testing System Software for IP-based Services in Mobile Networks”, IEEE SoftCOM 2007, Split-Dubrovnik, Croatia, September 27-29, 2007.
- Toni Janevski, **Aleksandar Tudzarov**, Perivoje Stojanovski, Dusko Temkov, “Applicative Solution for Easy Introduction of WLAN as Value-added Service in Mobile Networks”, IEEE VTC 2007 – spring, Dublin, Ireland, 22-25 April 2007.
- Toni Janevski, **Aleksandar Tudzarov**, et. al., “Unified Billing System Solution for Interworking of Mobile Networks and Wireless LANs”, The 13th IEEE Mediterranean Electro-technical Conference - MELECON’06, Malaga, Spain, May 16-19, 2006.
- Toni Janevski, **Aleksandar Tudzarov**, et. al., “Integrated AAA System for PLMN-WLAN Interworking”, TELSIKS 2005, pp.352-355, Nis, Serbia and Montenegro, September 28-30, 2005.
- Toni Janevski, **Aleksandar Tudzarov**, et. al., “Applicative Solution of a Billing System for Interworking of Mobile Networks and Wireless LANs”, IEEE SoftCom 2005, Split, Croatia, September 15-17, 2005.

mb!
MoreBooks! publishing
yes
i want morebooks!