

Protocols and Algorithms for the Next Generation 5G Mobile Systems

Aleksandar Tudzarov

T-Mobile Macedonia, Orce Nikolov bb, 1000 Skopje, Republic of Macedonia

Email: aleksandar.tudzarov@telekom.mk

Toni Janevski

Ss. Cyril and Methodius University, Faculty of Electrical Engineering and Information
Technologies, Karpos 2 bb, 1000 Skopje, Macedonia

Email: tonij@feit.ukim.edu.mk

Received: March 20, 2011 Accepted: April 20, 2011 DOI: 10.5296/npa.v3i1.656

Abstract

In this paper we propose new design framework for network architecture of future 5G mobile networks, including protocols and algorithms. The proposed design is based on user-centric mobile environment with many wireless and mobile technologies on the ground. In heterogeneous wireless environment changes in all existing wireless technologies is not possible, so each solution towards the next generation mobile and wireless networks should be implemented in the service stratum, while the radio access technologies belong to the transport stratum regarding the Next Generation Networks approach. In the proposed design the user terminal has possibility to change the Radio Access Technology - RAT based on certain criteria. For the purpose of transparent change of the RATs by the mobile terminal, we introduce so-called Policy-Router as node in the core network, which establishes IP tunnels to the mobile terminal via different available RATs to the terminal. The selection of the RAT is performed by the mobile terminal by using the proposed user agent for multi-criteria decision making based on the experience from the performance measurements performed by the mobile terminal. For the process of performance measurements we introduce the QoS PRO procedure for control information exchange between the mobile terminal and the Policy Router.

Keywords: 5G, Heterogeneous Wireless Networks, Quality of Service (QoS), Policy Routing.

1. Introduction

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 [1]. So, we may have different Radio Access Technologies (RATs) today and new RATs in the future (e.g., LTE-Advanced), but the common technology 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, so there will be no circuit-switching as it existed in 3G systems, such as UMTS. On the other side there are a lot of efforts 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 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 the 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 [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. On the other side, in the IP world, the main principle from 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 necessary Quality of Service, as well as authentication, authorization, accounting and security functionalities.

In this paper we provide complete functional architecture for 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. Then, 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 fully functional the proposed 5G network architecture.

The remainder of this article is structured as follows. Section 2 covers the models for interoperability in heterogeneous wireless environment. Section 3 presents the design of the

network architecture for 5G mobile networks. In Section 4 we provide description of use-cases for the proposed 5G architecture. Radio Access Technology Selector is the subject in Chapter 5. The proposed system for performance measurement in next generation wireless networks is given in Section 6. Finally, Section 7 concludes the paper.

2. Interoperability in Heterogeneous Wireless Environment

The challenge in the design of the terminals is connected to 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 reconfigurable 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 also maximize the battery power [4]. As a result of growing level of acceptance of the wireless technologies in different fields, 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 in multiple-access radio resource management. In these hierarchical schemes, local resource managers of the various wireless technologies provide interaction with the central entity to perform a joint optimization of available resources. In parallel with the evolution of cellular data systems, we have evolution of WLAN, i.e., IEEE 802.11 networks. The 802.11 systems handovers between different AP (Access Points) 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 for handover.

Taking into consideration the existence of various wireless technologies evolved in parallel, the concept of emergence of heterogeneous networks is not new. In heterogeneous wireless networks the concept is "always best connected" (always associated with the best quality), aimed at client terminals. This approach leads to the emergence for 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, the mobile terminal 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 a 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 access a single segment which will place the connection with the application servers in and out of operator's network. To meet the relevant requirements of the user applications there are generally considered two possible models for interoperability between building blocks of radio access technologies within the heterogeneous system. First one refers to a centralized operator access network, while the second one defines the Internet model of interoperability. The first models involves introducing a certain level of integration between the radio access technologies, and in that direction have been made different analyses as well as different standards that define the architectures for realization of the vertical handover between different access technologies involved in the heterogeneous domain, [5], [6]. The introduction of this model implies protocol interoperability at lower levels of communication protocol model, i.e. physical and link levels. The second model is called the Internet model, which is a focus for further development in this paper, and it refers to provisioning of continuity of the customer service when there are independent radio access technologies available to the mobile terminal by inter-connecting on the network level, [7]. In this case, interoperability between network technologies is done on the network protocol level, 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 it 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 concerns the way in providing interoperability. Apart from this difference, very important are vertical handover between access technologies and the conditions or circumstances which trigger handovers. 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 operators who have cooperation. In such networks 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 interoperate regardless of the user's operators, which provide wireless access networks to the user. In these methods, generally speaking, vertical handover is accomplished as a result of measurements of the main Quality of Service (QoS) parameters or service experience to the user, [8].

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 all of them 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 ensure the best possible service quality for a satisfactory price certainly appears the need for parallel usage of the variety of access technologies in order to realize the user requirements, [4]. The heterogeneity of these networks allows the user terminal to perform a selection of radio access technologies depending on given preferences, [9]. This choice provides better conditions for user applications regarding the given choices.

3. Design of 5G Network Architecture

Figure 1 shows a system model that proposes design for a 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 an 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 to have all of them active at the same time, with aim to have this architecture to be functional.

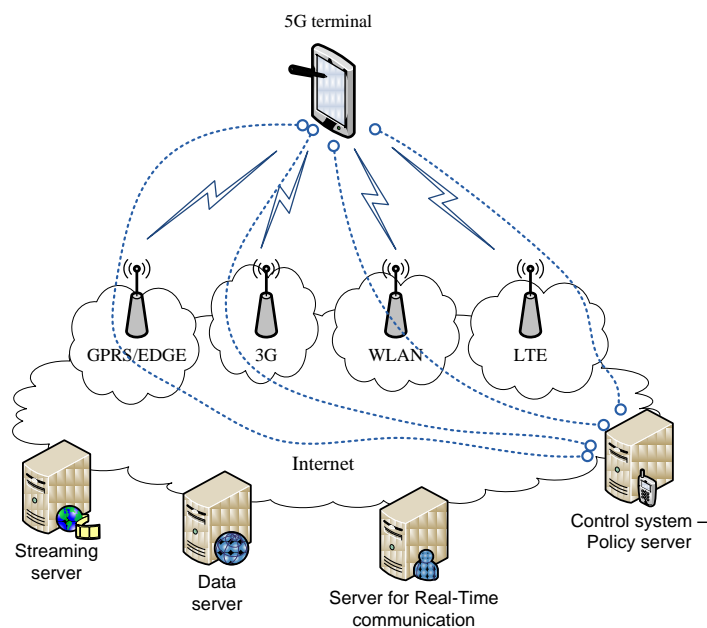


Figure 1. Functional architecture for 5G mobile networks

The first two OSI levels (data-link level and physical level) are defining the radio access technologies through which is provided access to the Internet with more or less QoS support mechanisms. However, this is further dependent upon the radio access technology (e.g., 3G and WiMAX have explicit QoS support, while WLAN has not). Then, 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 the 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 is a unified and 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. Considering that, the establishment of communication from end to end between the client and 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 for 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 Internet connection end-to-end, 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 uniquely and 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. Then, change of any of the parameters of the socket means change of the socket, that is, closing the socket and opening a new one. This means, ending the connection and starting a new one. This approach is not-flexible, and it is based on today's Internet communication.

In order to solve this deficiency we propose a new level that will take care of the abstraction levels 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 it can determine the quality of service for each transmission technology. It 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 picture regarding the quality that can be expected from applications of the user 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.

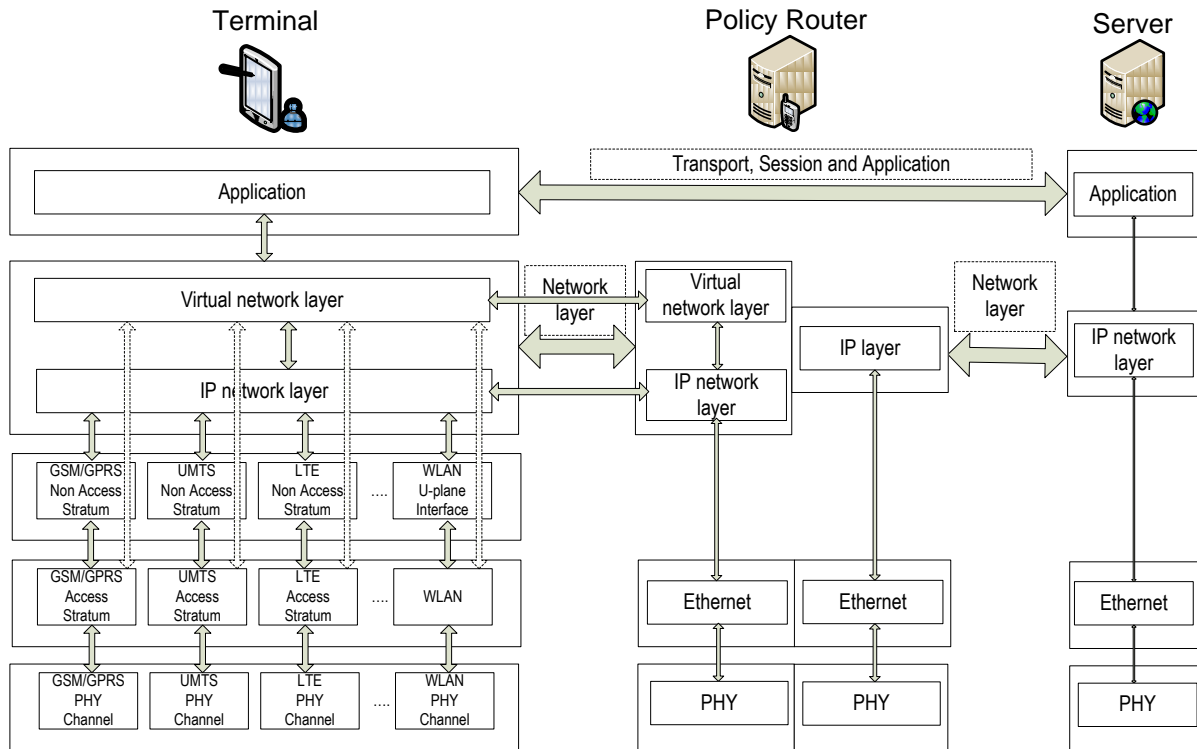


Figure 2. Protocol layout for the elements of the proposed architecture

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 named here as 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. The way IP packets are routed through tunnels, or choosing the right tunnel, would be served 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 routing based on the policies, are 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).

4. Description of use-cases in the proposed network architecture

Heterogeneity of wireless networks enables the user terminal to perform a selection of

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 some 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 logically is divided into several cooperative software modules which perform different functionalities. In Figure 3 is given block-diagram of the software modules in the virtual network layer.

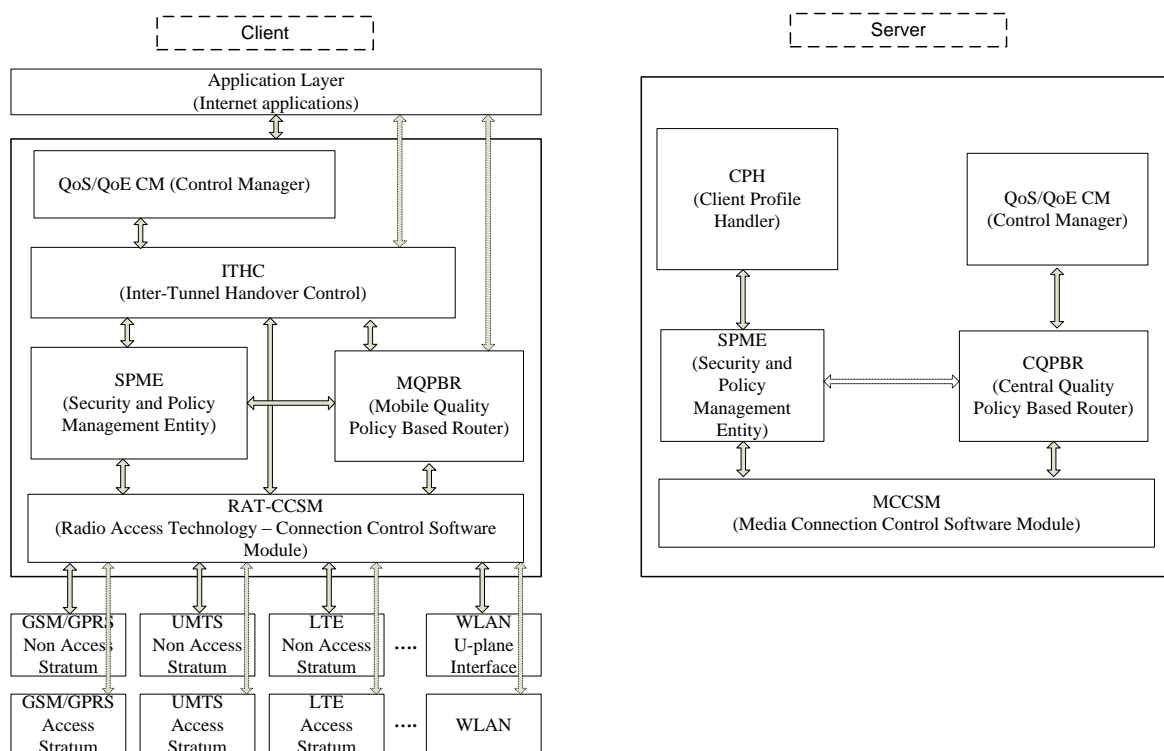


Figure 3 Software diagram of the proposed virtual network layer

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

- 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 its determined position within the architecture for 5G heterogeneous systems. Providing functionality between software modules is done through precisely defined interfaces to other modules and with appropriate links between peer protocol modules on both sides of the architecture.

4.1 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.

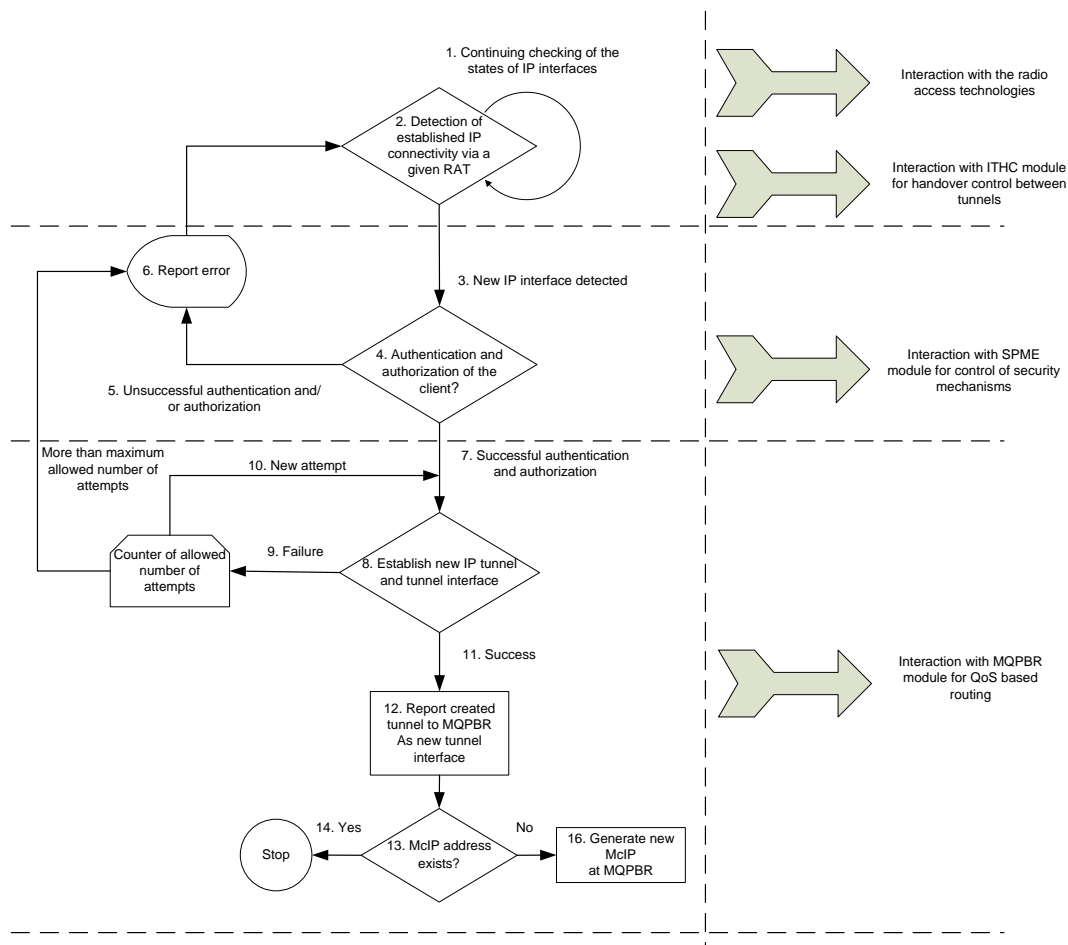


Figure 4. Block diagram for establishment of Internet tunnel for a given radio access technology

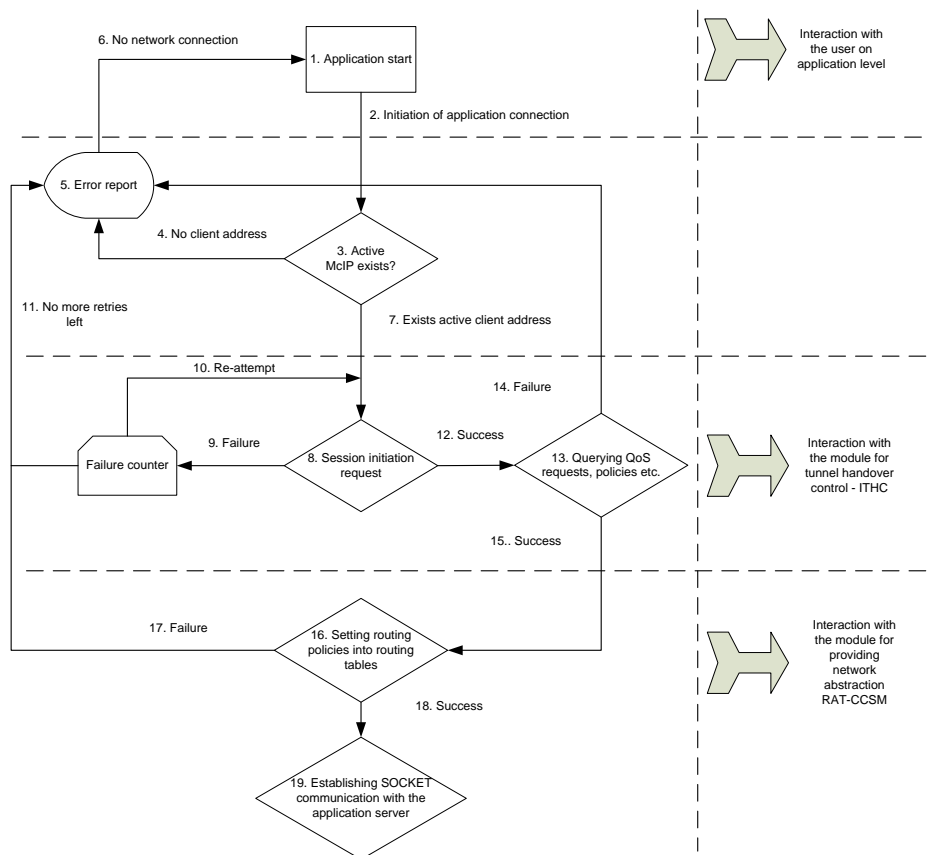


Figure 5. Block diagram for establishing customer service

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), starts mechanism of recurrence of the process that repeats via the steps of 8, 9, 10, retrying until it exceeds the allowable number of attempts defined in the algorithm. 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.

4.2 Establishment of customer service

The process of establishing of the customer service starts with 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 in Figure 5. The applications tries to establish a session, which

positive outcome results in querying the MQPBR for QoS request for the given application and policies regarding the user preferences, application (i.e. service) preferences, as well as access networks preferences. The establishment of customer service from the mobile terminal is outlined in the diagram shown in Figure 5.

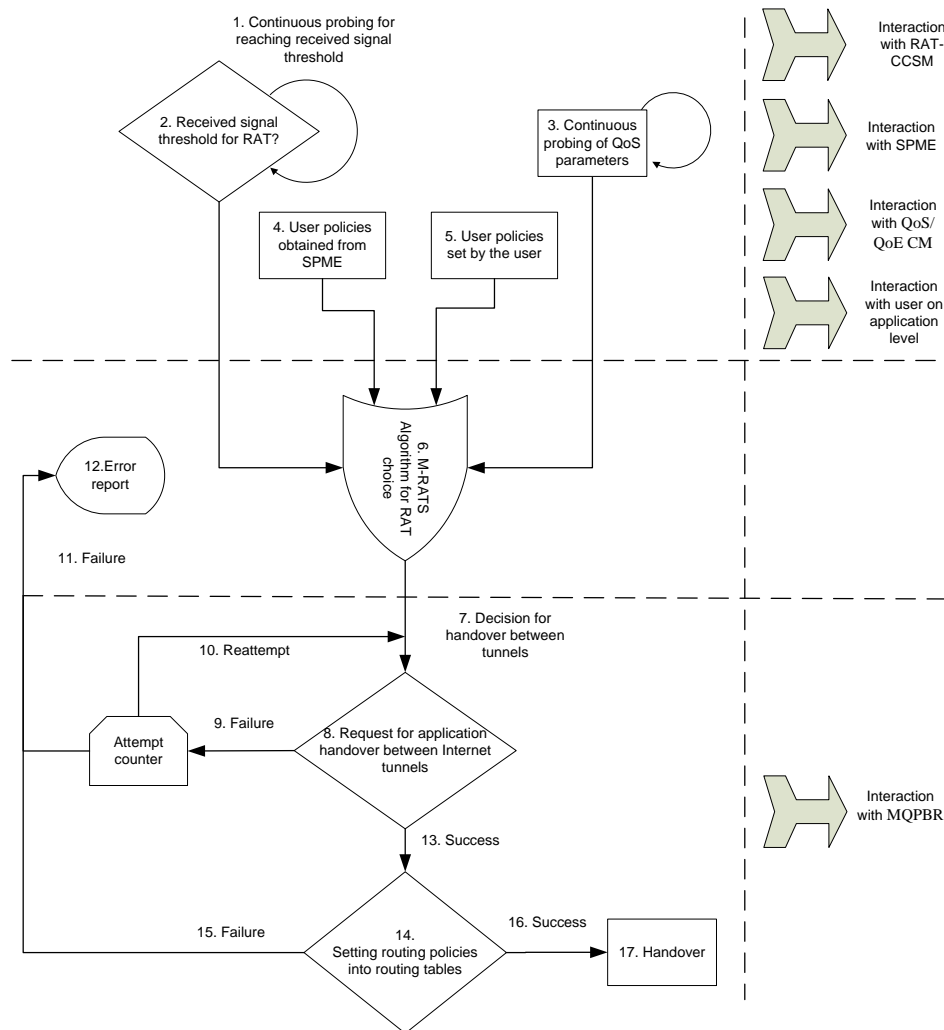


Figure 6. Block diagram for changing RAT – handover between IP tunnels

4.3 Process for changing Radio Access Technology - tunnel handover

The process of changing radio access technology begins with the phase of collecting information as an input for the handover algorithm. This activity is covered in steps 1, 2, 3, 4 and 5 of the diagram in Figure 6, which are 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 following phase of this process consists of an analysis of the input parameters through an intelligent algorithm (step 6), with aim to decide on handover. This procedure is repeated for each user session. The algorithm used for decision making mechanism is outlined in the following section.

5. Radio Access Technology Selector

In this section we provide description of a novel algorithm for radio networks selection in heterogeneous environment, which is created using biologically inspired algorithms. The algorithm consists of four building components as shown in Figure 7. First component or module is a set of parallel Fuzzy Logic (FL) controllers, which has as an input the measurements data for different selection criteria, including user requirements, QoS requirements, service policies, as well as radio link conditions in different wireless technologies available to the mobile user. The second module is multi-criteria decision mechanism algorithm, which uses as inputs the outputs of the FL controllers from the first module. 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) [10], [11], mechanism which dynamically modifies the functions of FL controllers in the first module (shown in Figure 7).

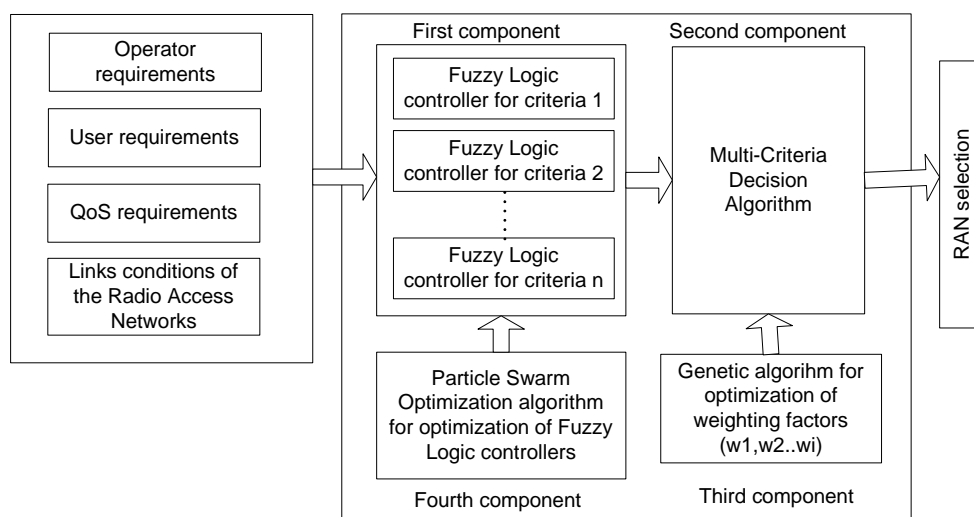


Figure 7. Radio Access Network selection scheme

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 paper 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.

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, such measured parameters from each network are received signal power and signal to noise ratio on link layer. 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. The proposed scheme assumes that

mobile terminal has enough processing power, memory capacity and battery support, so it can provide functionalities described above. Considering the Moore's law for computers, where we can today place mobile terminals as well, one may assume that near year 2020 mobile phones will have processing capabilities of today's power computers. So, the mobile phone can operate a database, do processing on given time intervals using measurements data from a given timeframe in the past.

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 8, best buffer size can be defined in borders of 300- 400 history intervals.

History intervals, as shown on Figure 8, 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.

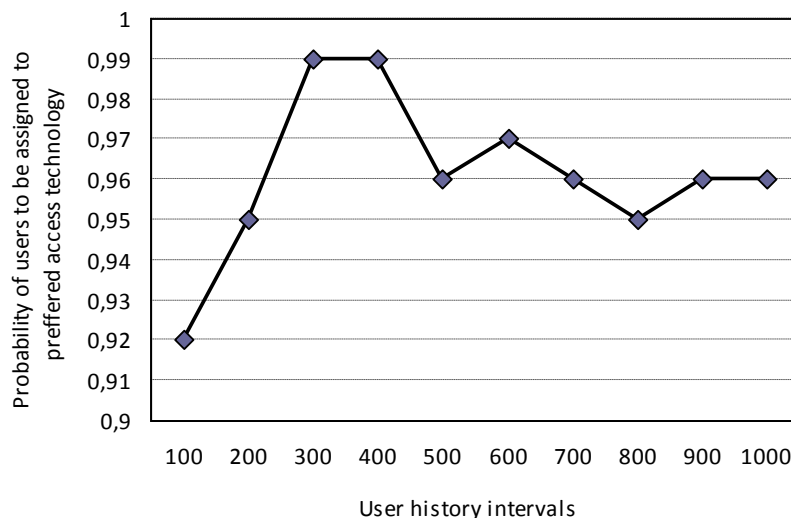


Figure 8. Initial buffer size dependency

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:

$$v_i = v_{i-1} * C_v + \sqrt{1 - C_v^2} * v_{mean} * N \quad (1)$$

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:

$$C_v = \left(\frac{-dt * a_{mean}}{v_{mean}} \right) \quad (2)$$

where N is Rayleigh distributed magnitude with mean 1 and a random direction. Parameter v_{mean} is the mean speed of mobiles, and v_{mean} was set to 10 km/h, a_{mean} has been set to 1 km/h², which are typical values for urban environment. Figure 9 shows the users, in the marked spots, over the simulated environment. Four types of services are considered in the simulation and they are equally distributed among the users, the voice calls, the low bit rate real-time video telephony, the high bit rate video, and the non-real-time data traffic.

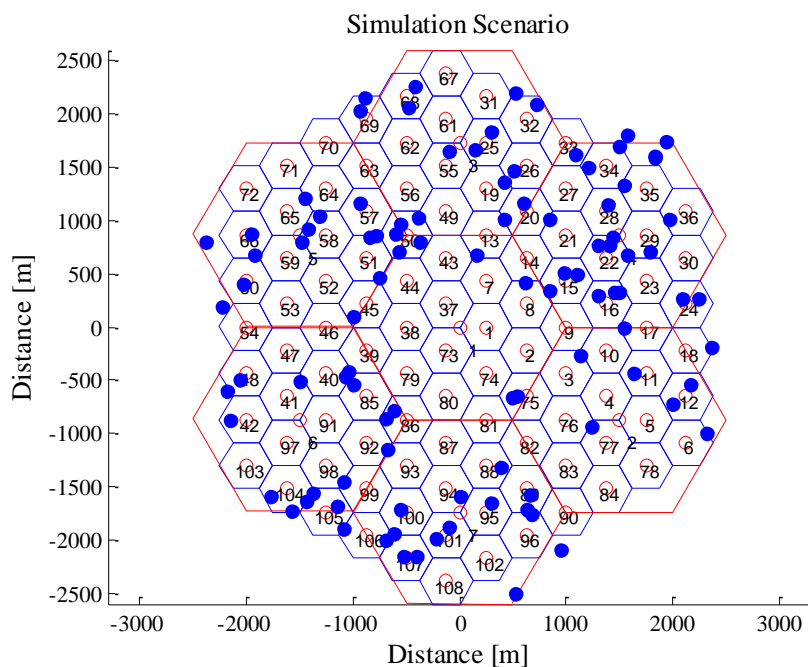


Figure 9. Simulation Scenario

Further we need to model the four types of services used in the simulation analysis in the following part of this paper. 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:

$$[service_latency (ms), service_bandwidth (kbps)] \\ \{ [100, 64]; [200, 128]; [400, 256]; [800, 512] \} \quad (3)$$

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. 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. For the purpose of simulation Fuzzy logic controllers were designed to fit the FL controller parallel scheme given in Figure 7. During this design process a Particle Swarm Optimization (PSO) algorithm is used 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. Outputs from the FL controller present so-called degree of membership of each RAT scaled to input variables regarding Fuzzy Logic rules.

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 [10]. 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 for 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.

We have compared in Figure 10 the proposed algorithm with almost all other relevant algorithms. 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).

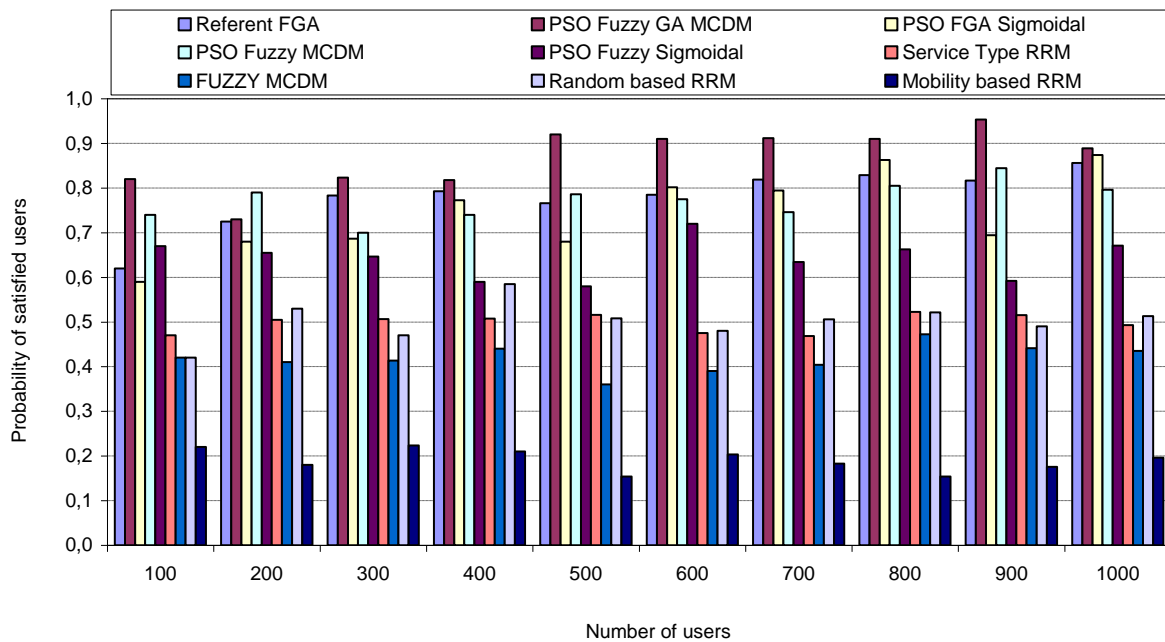


Figure 10. Probability of satisfied users

In the Figure 10 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. The proposed wireless access selection algorithm in this paper achieves around 40% enhancement compared to service-based and random-based selection algorithms. Also, 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.

In the following section we provide a novel concept for performance measurements, which should accompany the presented radio access technology intelligent selection algorithm, by providing mechanism for feeding with historical data the mobile terminal intelligent RAT selector, because single user can have experience only from its own behavior, while dedicated network node(s) can store measurements data from different users, different services and different radio access networks. Performance measurements are considered as very important in recommendations for the Next Generation Networks (NGN) in access, core and transit networks.

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

NGN consist of support functionalities for data transport, and control transport, as well as functionalities for support of services and applications. The measurement of traffic is a basic control activity in order to provide Quality of Service, [12]. So, performance measurement is an intrinsic component in NGN, and it is usually performed at edge (border) network nodes, in access, core and transit networks. Also, it can be performed by the mobile terminals in the

wireless environment, and here this is very important to create mobile user assisted probing of the RAT performances, [13], [14].

Figure 11 presents a view of the general architecture for Management for Performance Measurements (MPM) in NGN, with added functionalities on the side of the mobile terminal (CPE – Customer Premises Equipment).

The network architecture consists of the following entities:

- Performance Measurement Execution Functional Entity (PME-FE) is an entity to perform performance measurements. PME-FE is responsible for three groups of functionalities: performance measurements, process measurements and single configuration entity conducting the measurements;
- Performance Measurement Processing Functional Entity (PMP-FE) is an entity for processing the measurements. PMP-FE is responsible for two groups of functions: processing of measurements and configuration of measurement tests across the network architecture. Functions for processing of measurements include collecting measurement reports, their analysis, aggregation as well as analysis of measurements in cyclical periods; and
- Performance Measurement Reporting Functional Entity (PMR-FE) is an entity reporting the performed 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.

The idea is to perform continuous performance measurements by the customer premises equipment (CPE) without introducing additional probing systems and this to be used towards the management of the performance as an aid in decision-making on resource allocation in next generation networks, respectively, and thus directly affect the provisioning of the required QoS for each of the realized services that customers use. Figure 11 shows the possible configuration for performance measurements using RTP / RTCP. On this Figure is presented Mu interface to Resource and Admission Control Function (RACF), which is an essential part of the process of decision-making performance in the allocation of resources. In the proposed architecture in this paper the interface for measurement data is between the QoS / QoE CM software module and software module for control of handovers between wireless technologies (ITHC).

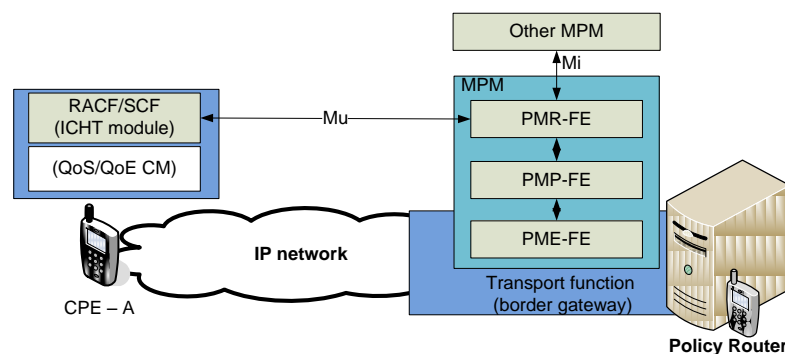


Figure 11 Network performance measurements using RTP / RTCP and RTCP extension, based on NGN [12]

The procedure for notifications based on qualitative performance measurements obtained by RTP / RTCP-based scheme is shown in Figure 12.

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.

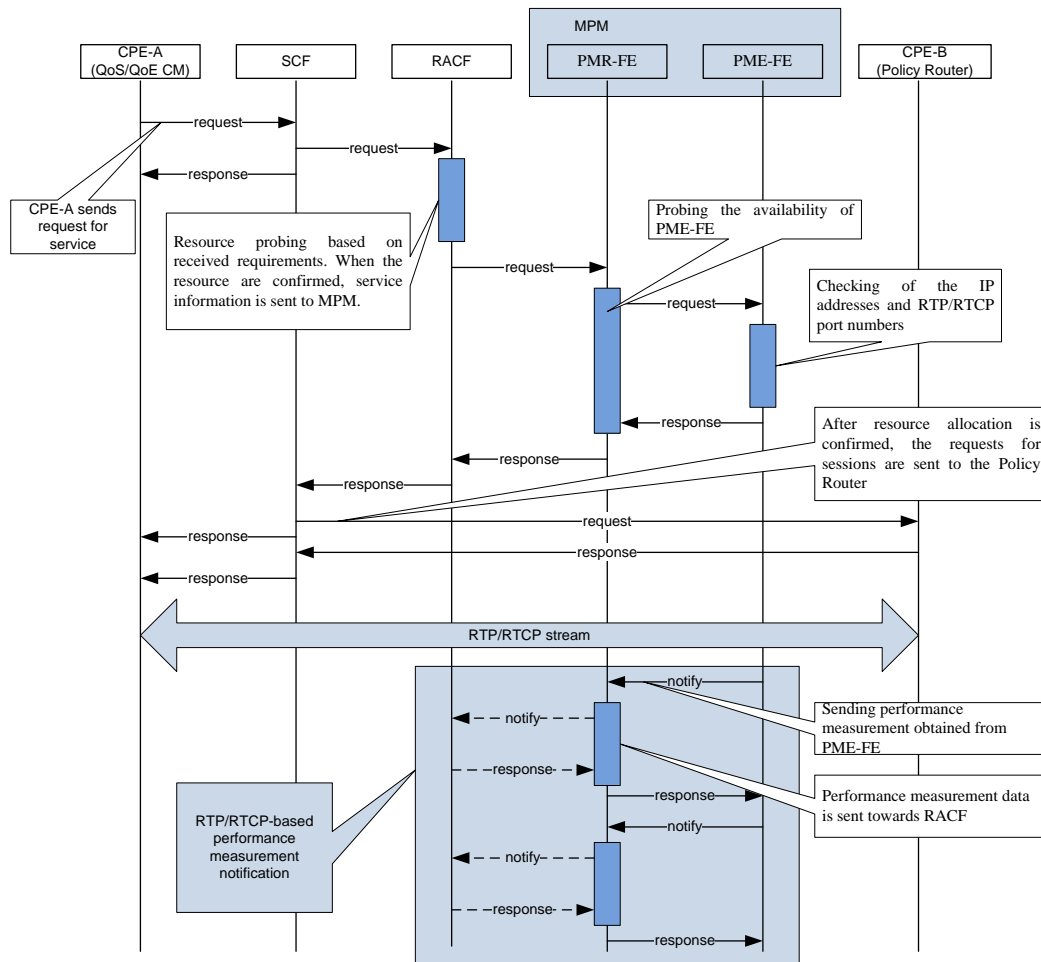


Figure 12. RTP / RTCP-based diagram for performance measurements

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

- 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, which is achieved through connection to the client with the IP world; and
- 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 in itself all the essential and necessary parameters by which the exchange between the control and measurement entities, in particular between MQPBR and CQPBR software modules in the proposed architecture, would enable the mutual synchronization in order to achieve proper operation. Protocol procedures that should be supported by QoSPRO protocol are shown in Figure 13. Decisions for change of the radio access technology are executed 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.

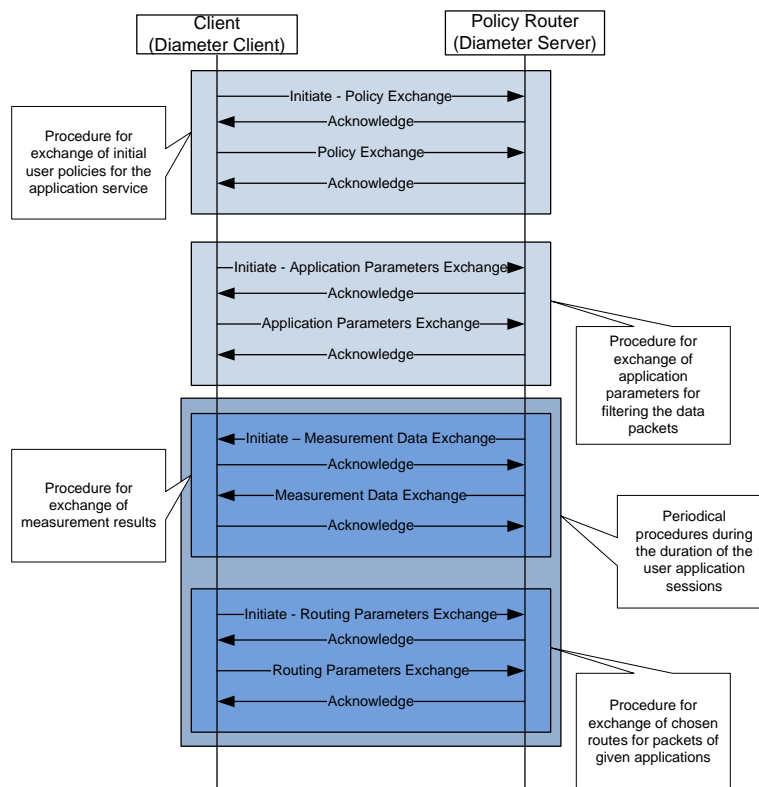


Figure 13. QoSPRO – novel procedure for the information exchange in the proposed architecture

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

7. Conclusion

The development of the mobile and wireless networks is going 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, as well as there is migration of all services in IP environment, including the traditional telephony and even

television, besides the traditional Internet services, such as web and electronic mail as most used among the others. On the other side, mobile terminals are obtaining each year more processing power, more memory on board, and longer battery life for the same applications (services). It is expected that the initial Internet philosophy of keeping the network 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 paper we have defined completely novel 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, which shall be 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 for that purpose in this paper, called Quality of Service Policy based ROuting (QoS PRO). Finally, the proposed RAT selector algorithm, based on genetic algorithms and particle-swarm optimization, provides highest probability of satisfied users regarding their requirements from the access network.

The proposed architecture for future 5G mobile networks can be implemented using components of the shelf (existing and standardized Internet technologies) and its implementation is transparent to the radio access technologies, which makes it very likeable solution for the next generation mobile and wireless networks.

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] T. Janevski, "5G Mobile Phone Concept", IEEE CCNC 2009, Las Vegas, USA, 10-13 January 2009. <http://dx.doi.org/10.1109/CCNC.2009.4784727>
- [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. <http://dx.doi.org/10.1016/j.comcom.2008.01.044>
- [6] W. Luo, E. Bodanese, "Optimising Radio Access in a Heterogeneous Wireless Network Environment", IEEE International Conference on Communications, Dresden, Germany, 14-18 June 2009. <http://dx.doi.org/10.1109/ICC.2009.5199291>
- [7] Ha Nguyen Tran, M. 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. <http://dx.doi.org/10.1109/PIMRC.2009.5450191>

- [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. <http://dx.doi.org/10.1109/ICC.2007.932>
- [9] A. Tudzarov, T. Janevski, “M-RATS: Mobile-based Radio Access Technology Selector for Heterogeneous Wireless Environment”, Telfor Journal Vol.3 No.1 (2011).
- [10] 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. <http://dx.doi.org/10.1109/4235.985692>
- [11] J. Perez-Romero, O. Sallent, R. Agusti, and M. A. Diaz-Guerra, “Radio Resource Management Strategies in UMTS”, John Wiley & Sons, New York, NY, USA, 2005.
- [12] ITU-T, Y.2173, “Management of performance measurement for NGN”, September 2008.
- [13] 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. <http://dx.doi.org/10.1109/EURCON.2009.5167860>
- [14] 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. <http://dx.doi.org/10.1109/SOFTCOM.2007.4446130>

Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).