Real-Time Emulation of Heterogeneous Wireless Networks with End-to-Edge Quality of Service Guarantees: The AROMA Testbed

Miguel López-Benítez, Francisco Bernardo, Nemanja Vučević and Anna Umbert

Department of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC)
C/Jordi Girona 1-3, 08034 Barcelona, Spain

E-mail: [miguel.lopez, fbernardo, vucevic, annau]@tsc.upc.edu
Phone: +34 934015948
Fax: +34 934017200

ABSTRACT

This work presents and describes the real-time testbed for all-IP Beyond 3G (B3G) heterogeneous wireless networks that has been developed in the framework of the European IST AROMA project. The main objective of the AROMA testbed is to provide a highly accurate and realistic framework where the performance of algorithms, policies, protocols, services and applications for a complete heterogeneous wireless network can be fully assessed and evaluated before bringing them to a real system. The complexity of the interaction between all-IP B3G systems and user applications, while dealing with the Quality of Service (QoS) concept, motivates the development of this kind of emulation platform where different solutions can be tested in realistic conditions that could not be achieved by means of simple off-line simulations. This work provides an in-depth description of the AROMA testbed, emphasizing many interesting implementation details and lessons learned during the development of the tool that may result helpful to other researchers and system engineers in the development of similar emulation platforms. Several case studies are also presented in order to illustrate the full potential and capabilities of the presented emulation platform.

KEYWORDS

All-IP networks, beyond 3G, common radio resource management, end-to-edge quality of service, heterogeneous wireless systems, performance evaluation, real-time testbed.

1. INTRODUCTION

Since their appearance, mobile communication systems have been experiencing a constant evolution in order to support innovative services, provide increased data-rates and offer enhanced capabilities. As a result, a large number of wireless technologies such as Second/Third Generation (2G/3G) cellular networks, Wireless Local/Personal Area Networks (WLAN/WPAN) or broadcast networks are deployed nowadays. Some examples of the most popular technologies include the Global System for Mobile communications (GSM) [1][2], General Packet Radio Service (GPRS) [2], Enhanced Data rates for GSM/Global Evolution (EDGE) [2], Universal Mobile Telecommunications System (UMTS) [3], High Speed Packet Access (HSPA) [4][5], Long Term Evolution (LTE) [5][6], Worldwide Interoperability for Microwave Access (WiMAX) [7], Ultra Wide Band (UWB) [8], Digital Video Broadcasting (DVB) [9], IEEE 802.11 [10], Bluetooth [11], and so forth. These technologies were developed for different application scenarios and are therefore characterized by different capabilities in terms of capacity, transmission rates, coverage and cost. The complementary characteristics offered by different wireless technologies make it possible to exploit the trunking gain leading to a higher overall performance than the aggregated performances of the standalone networks [12]. Thus, trends in mobile and wireless communications are currently evolving towards the integration and joint management of different wireless access networks and technologies into heterogeneous wireless infrastructures, also referred to as Beyond 3G (B3G) networks. This concept assumes that different Radio Access Networks (RANs) can be cooperating components through which network providers can satisfy the wide variety of demands in a more efficient manner. The interconnection of different RANs is accomplished by means of a common Core Network (CN), which is usually based on the Internet Protocol (IP). The IP technology is becoming the cornerstone around which wireless technologies are converging. In this context, the concept of all-IP is commonly used to refer to those systems that provide IP-based multimedia services over IP-based transport along the whole network, i.e. through both the RAN and CN parts.

One of the main challenges that all-IP B3G heterogeneous wireless systems must overcome is the seamless interoperability and efficient management of different RANs, which is required in order to provide continuous and ubiquitous connectivity through different wireless technologies while preserving the negotiated Quality of Service (QoS) level for the end-user during the entire session lifetime. To this end, appropriate solutions are required for both the RAN and CN parts of the network. On one hand, efficient Radio Resource Management and Common Radio Resource Management (RRM/CRRM) strategies need to be developed to ensure an efficient and coordinated use of the pool of available radio resources provided by each one of the individual RANs. On the other hand, the CN features need also to be taken into account and efficient end-to-edge (e2e) QoS policies to coordinate the RAN and CN parts must be defined in order to provide the required e2e QoS levels.

Such strategies and policies for mobile communication systems have usually been evaluated by means of off-line system-level simulators. The use of simulation tools is common within the research and industrial communities and can be useful for obtaining preliminary results. Nevertheless, to conduct meaningful and appropriate studies, and to accurately assess the performance of innovative solutions before considering a prototype or full-scale deployment, the evaluation over realistic platforms is becoming essential as a step forward toward the implementation in a real system. Real-time emulators allow reproducing realistic scenarios to test algorithms, policies, protocols, services and applications under realistic conditions, thus constituting a powerful tool for evaluating not only the QoS but also the Quality of Experience (QoE) of the end-user, which could not be achieved by means of off-line simulations.

In this context, this work provides an in-depth description of a sophisticated real-time testbed for all-IP B3G heterogeneous wireless networks that has been developed in the framework of the European IST AROMA project [13]. This work emphasizes many interesting implementation details and lessons learned during the development of the tool that may result helpful to other researchers and system engineers in the development of similar emulation platforms. Several case studies are also presented in order to illustrate the full potential and capabilities of the presented tool. Some parts of the material presented in this work have already been published in referred conferences on the field of mobile wireless communications [14][15][16]. However, such publications either provided a very general overview or treated particular testbed aspects and capabilities independently, and they did not provide a global and comprehensive vision of the presented tool. This paper provides a detailed unified description of the whole developed platform, its full potential and its applicability. The aim of this work is to provide a holistic vision and discussion of the AROMA testbed and to show how this kind of emulation platform can be useful to carry out a wide range of accurate multidisciplinary studies.

The rest of this work is organized as follows. First, Section 2 provides a brief revision of previous related work and describes the motivation for developing the new testbed presented in this paper. Afterwards, Section 3 discusses its applicability. Section 4 provides a general overview of the testbed architecture, with more detailed descriptions of the three main parts of the testbed, namely the RAN, the CN and the e2e framework, being provided in Sections 5, 6 and 7, respectively. Several case studies illustrating the capabilities of the emulation platform are then presented in Section 8. Finally, Section 9 provides concluding remarks.

2. PREVIOUS WORK AND MOTIVATION

Different alternatives can be considered when evaluating algorithms, policies, protocols, services or applications for mobile communication systems. A first possibility is to use analytical models, which constitute a powerful tool for establishing, by means of mathematical formulas, a direct and clear relation between several system parameters and the system performance. However, the reliability and accuracy of analytical models is reduced by the significant assumptions and simplifications that are usually needed for the problem to be analytically treatable. Moreover, obtaining a closed solution may become in some cases extraordinarily difficult, if not impossible, especially as the complexity of the system under study increases. The increasing complexity of mobile communication systems hinders the study of a complete B3G heterogeneous wireless network by means of analytical models.

A second option is to implement the solution under evaluation in a real operating network or prototype and perform measurements in order to assess its performance. In this case, all the factors of the real network influence the obtained results. Hence, the performance can be assessed in an objective and reliable manner. Nevertheless, the deployment of a real network for experimentation purposes results prohibitively complex and economically unfeasible. Moreover, testing in existing production and commercial networks is typically forbidden since it presents a high degree of risk factors for service availability that network providers are not willing to assume.

An additional alternative, somewhere in the middle between the two previous approaches, is computer simulation. The system components and their behavior are modeled in software that runs in a computer. In general, this method requires an important amount of execution time to provide results, depending on the detail of the simulation model. Some assumptions and simplifications are required as in the case of analytical studies, but their amount normally is significantly lower, which implies a higher accuracy of results. Although the obtained results are not as realistic as in the case of real network or prototype experiments, simulation provides an acceptable degree of realism at a much reduced cost. Therefore, simulation represents a reasonable trade off among time, cost, accuracy and complexity.

Network research has successfully been relying on a combination of analytical models, computer simulation and network prototypes. All of them complement each other. While analytical modeling and simulations simplify some parts of a real environment in order to understand the impact of other factors, real world experiments aim at capturing the full interaction between all the involved parts. This interaction is difficult to model or simulate as it requires such a detail that it is not feasible. A powerful hybrid solution that combines the previous approaches is referred to as *emulation*, which merges the simulation of some parts (making use of analytical models in some cases), and the real implementation and live running of other parts.

As the complexity of mobile communication systems has been increasing, more sophisticated and detailed evaluation platforms have been required. As a result, a large number of evaluation tools have been documented in the literature for wireless communication systems in general and heterogeneous wireless networks in particular. In the domain of network research, many simulation tools have been presented and several emulation and real network testbeds have been proposed in the past.

The Bay Area Research Wireless Access Network (BARWAN) [17] constituted one of the first heterogeneous wireless research testbeds. It was implemented in the metropolitan area of San Francisco in 1996 and integrated different range wireless networks (regional, metropolitan and in-building). To some extent, the work in the BARWAN project established the foundations of subsequent research in the ambit of testbeds for heterogeneous wireless access networks.

Thenceforth, a wide variety of research testbeds have been implemented to investigate various aspects of wireless communication networks, from the fixed network part, based on optical networks (e.g., NCIT*net 2 [18] and CREATE-NET [19] testbeds) or generic packet-switched networks (Emulab [20]), to the radio access part based on single wireless technologies such as WiMAX [21][22], Multiple-Input-Multiple-Output (MIMO), and software defined radio (UCLA Hybrid Network Testbed [23]).

Some other examples of research testbeds related to heterogeneous wireless networks include the testbed developed during the Mobility and Differentiated Services in a Future IP Network (Moby Dick) IST project, where an isolated testbed integrating access networks based on IEEE 802.11b and UMTS technologies was implemented. Unfortunately, a complete vision of a 3G operator was not given [24]. Afterwards, the DAIDALOS (Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimised personal Services) and DAIDALOS II IST projects [25] continued the work started in Moby Dick and developed an open architecture based on a common network protocol (IPv6), mainly to study mobility issues in a B3G heterogeneous network. The Berlin's Beyond-3G Testbed and Serviceware Framework for Advanced Mobile Solutions (BIB3R) solved the mobile operator's network interconnection and supported additional access technologies like pre-WiMAX and Flash-OFDM [26]. Another relevant heterogeneous wireless testbed was implemented in the WHYNET project [27]. This project developed a wireless hybrid network testbed to assess cross-layer interactions in heterogeneous wireless systems based on sensor and mesh networks.

The main drawback of most of the implemented tools is that they have often targeted specific and particular aspects of the overall research in heterogeneous wireless networks, depending on the purpose for which the tool was developed. The modeling detail of these tools strongly depends on the type of work being conducted. Therefore they are frequently specific-purpose tools with significant oversimplifications with respect to a complete real system. Although each existing platform is valid within its specific research framework, we believe that besides the important but often isolated simulators and specific targeted research testbeds, it is also necessary to provide researchers with an environment that reflects a complete real world heterogeneous wireless network.

In this context, this work reports an ambitious and sophisticated testbed that enables the real-time emulation at the packet/slot level of an entire e2e all-IP heterogeneous wireless network with a high level of implementation detail. The complexity of the interaction between all-IP B3G systems and user applications, while dealing with the QoS/QoE concept, motivates the development of such complex platform where different solutions can be accurately tested under realistic conditions that could not be achieved by means of simplified specific-purpose tools.

3. APPLICABILITY

The versatile tool presented and described in this work is a suitable and powerful platform for a broad variety of multidisciplinary studies ranging from specific low-level studies related to individual sub-components of the network and its associated algorithms and protocols, to e2e testing of services and user applications at the application level. Several live demonstrations performed in international conferences have already demonstrated the applicability of the presented tool [28][29]. The following lines provide some examples.

Regarding the RAN part of the testbed, the studies that can be carried out embrace the validation and performance evaluation of advanced RRM and cross-layer RRM algorithms for the specific technologies implemented in the testbed (see e.g. [30]), including admission control, congestion control, power control, radio resource allocation, handover management, and transmission parameters management, etc. Another kind of studies related to the RAN part comprise the validation and performance evaluation of CRRM-related issues (see e.g. [31][32]), including CRRM architectures, inter-RAN communication mechanisms, RAN selection and load-sharing algorithms, and Vertical Handover (VHO) execution and coordination policies, among others. The development, assessment and optimization of mechanisms allowing an automated tuning and self-optimization of the RRM/CRRM algorithms and their corresponding parameters are also possible.

Concerning the CN part, the studies that can be performed include the analysis and evaluation of QoS and traffic engineering technologies and their interactions under an all-IP network, as well as the evaluation of different protocol stacks and security mechanisms, etc.

The testbed can also be employed to evaluate different e2e aspects such as the variation in the final QoS experienced by a user running real IP-based multimedia applications [33] when changing different QoS management policies and algorithms affecting the CRRM functionalities, the IP RAN, the CN and the overall system. Therefore, the testbed can be useful to determine appropriate network configurations needed to satisfy the QoS requirements for selected applications, to assess the impairments in QoS perception related to specific network conditions or inter-system changes, and so forth. Notice that conventional off-line system simulators are not well suited for investigating real-time services and multimedia applications on dynamically varying radio environments. Certain effects such as end-user perception, e2e QoS, QoE or human interactions with the network (i.e., how the user reacts to the network performance), cannot be fully understood with off-line simulation results or analytical studies, and are more appropriately studied with real-time emulators. Other e2e issues such as QoS-aware mobility mechanisms [34] and signaling procedures [35] (session negotiation, session establishment, inter-system handover, session re-negotiation, session dropping and session closing) can be evaluated and optimized with the testbed as well.

In general, the testbed models a comprehensive network that can be easily configured to reflect a wide range of scenarios, thus providing researchers with a powerful tool for proof of concepts and allowing system engineers to validate and optimize their designs and algorithms before bringing them to a real system.

The different kinds of studies listed above have usually been tackled by means of specific-purpose evaluation tools particularly envisaged for the concrete study being carried out. In this context, the aim of this work is to present the main design and implementation approaches that allowed us to develop a comprehensive and complex emulation framework able to face all the previous network research problems with a single evaluation platform.

4. TESTBED ARCHITECTURE

4.1. General Description

The AROMA testbed allows the real-time emulation of an all-IP heterogeneous wireless network composed of the UMTS Terrestrial Radio Access Network (UTRAN) with High Speed Downlink/Uplink Packet Access (HSDPA/HSUPA) Release-6, GSM/EDGE Radio Access Network (GERAN), and Wireless Local Area Network (WLAN) as well as the corresponding common CN based on DiffServ technology [36] and Multi-Protocol Label Switching (MPLS) [37]. The evaluation platform emulates, in real-time, the conditions that the behavior of the all-IP heterogeneous network, including the effect of other users, produces on a given User Under Test (UUT), who is unique, when making use of real multimedia IP-based. This approach enables an accurate e2e evaluation of real user applications over a realistic and complete all-IP heterogeneous network with advanced RRM/CRRM algorithms and e2e QoS management policies.

4.2. Software and Hardware Platform

The AROMA testbed is composed of twenty off-the-shelf Personal Computers (PCs). Two of them (applications PCs) run the Windows Operating System (OS) while the other eighteen PCs run the Linux OS. An additional PC connected to the Internet is used as a firewall in order to enable restricted remote access to the testbed.

The Windows OS was selected for the applications PCs due to the wide availability of popular and easily configurable client/server applications for such OS. Any IP-based multimedia applications, including real world applications such as those provided in the

Internet (web browsing, video streaming, video conferencing, audio conferencing, voice over IP, gaming, and so on) can be installed in the applications PCs and run over the testbed in order to test the end-user QoS perception under specific scenarios or to determine the optimum network configuration for providing certain services. This property is one of the distinguishing features of the testbed with respect to other conventional evaluation platforms. Both commercial and open source applications can be installed and evaluated. Some of the applications employed in our studies are listed in Table 1. In order to evaluate the user QoS perception, the (degraded) multimedia contents received at the user side are captured and compared to the reference (original) contents stored in the sever, according to the QoS metrics recommended in [38][39][40][41]. The software employed to capture the user perception is also indicated in Table 1.

The Linux OS was selected for the other eighteen PCs of the testbed for its capability to assure appropriate levels of real-time management while guaranteeing a high degree of flexibility. The capabilities provided by Linux to interact at low level with the kernel offer the possibility to tune accurately the performance required by the testbed, especially in the issues related with the real-time execution and management.

To implement real-time operation a very high computational power is required. These computational requirements are out of the scope of today's off-the-shelf PCs. Therefore, a cluster of PCs has been constructed in order to distribute the computational load throughout different processors. To this end, a software tool named Communications Manager (CM) was developed. The CM offers the possibility to seamless distribute different software pieces across several machines with a network interface and run them in parallel in order to improve the overall performance or achieve a certain real-time execution constraint. The CM was designed and developed to make this distribution completely transparent to the software running on top. The CM may be understood as an abstraction layer between the hardware and/or OS (if any) and the running software. It hides any specific hardware- and/or OS-related aspect of the possibly heterogeneous platform compound of multiple machines and/or OSs. Software modules are unaware of the number of machines and OS actually being used. Software modules directly interface the CM regardless of the logical connections established among them.

Besides hiding the underlying overall platform, the CM is also in charge of other tasks such as providing a centralized method to start, stop, debug and monitor the whole running software, providing timing control of software modules in order to allow their coordinated real-time execution, and gathering multiple forms of data provided by the software modules in order to be displayed in real-time or post-possessed off-line after the testbed execution. Moreover, the CM solves the problem of integrating different software pieces produced by different programmers but developed following a common set of rules, since it provides a common Application Programming Interface (API) that eases the control of the execution cycle, the statistics collection method, and the inter-process communication among software modules running on different machines. The CM is written in plain C code in order to minimize the use of specific hardware- and/or OS-dependent functions, is POSIX compliant [42] and can be ported to any platform running a Unix-based OS. For more details about the CM, the reader is referred to [43].

Figure 1 shows all the entities and logical connections of the AROMA testbed. The physical connections of the underlying hardware platform are shown in Figure 2. Full line black connections in Figure 1 correspond to user data interfaces, whereas dashed blue and red connections correspond to control and e2e QoS signaling interfaces respectively. The UUT has at its disposal two stand-alone PCs: one PC (applications client) is used to run the UUT application, e.g. a commercial web browser or a streaming video player, while the other PC is used to run the main functionalities associated to the User Equipment (UE). This second PC provides a graphical interface for session control through which sessions can be activated, deactivated or modified in real-time during the testbed execution, specifying different QoS parameters (e.g., conversational, streaming or interactive QoS class, and requested throughput in uplink and downlink). This interface enables the human interaction with the testbed in real-time. To test symmetric services such as video conference and to serve multimedia applications such as web-browsing or streaming, a correspondent node (applications server) is run in a stand-alone PC.

The three implemented RANs are emulated using three PCs for UTRAN (uplink, downlink and common functionalities), one PC for GERAN and one PC for WLAN. The CN has been built using seven PCs acting as routers: three PCs serve as edge routers, two Ingress Routers (IRs) and one Egress Router (ER), and four PCs identified as Core Routers (CR) interconnect the edge routers following a typical unbalanced *fish topology*.

A Traffic Switch (TS) is used to establish different connection configurations between the RANs and the IRs in the CN. It captures the IP packets for the UUT in both downstream and upstream directions, passes them to the appropriate RAN (where the UUT is connected to) to make the real-time emulation and afterwards forwards them to either the UE (in downstream) or the IR where the RAN is supposed to be connected to (in upstream). For emulated users, the Traffic Generator (TG) PC is in charge of generating real IP traffic to load the CN according to the traffic amount that active users generate in the system. Obviously, the generation of this traffic is coordinated with the traffic emulated in the RAN part.

The Bandwidth Broker (BB) is in charge of coordinately managing the CN resources and the QoS mechanisms associated to the CN. Its counterpart for the RAN is the Wireless QoS Broker (WQB), which is responsible for the QoS management functions associated to the heterogeneous RAN. The CRRM functions, responsible for the efficient management of the pool of radio resources provided by each one of the individual RANs, are also implemented in the WQB PC. While the WQB and BB entities take QoS and resource management decisions for the RAN and CN domains respectively, the e2e decisions are taken by the Master Policy Decision Point (Master PDP) entity, which is also implemented in the WQB PC for simplicity reasons.

Finally, the Advanced Graphical Management Tool (AGMT) PC runs a software application that provides a graphical interface to easily configure the initialization parameters of each entity, control the execution flow of the testbed, collect logged data, obtain and display statistics during the real-time execution of the testbed (performance measurements as well as state information parameters), and save trace files for post-processing. The shaded yellow area in Figure 1 embraces all the machines controlled by the AGMT. The interested reader can find a more detailed and complete description of the AGMT and its use in the public documentation and demonstration videos provided in the AROMA project's web page [13].

4.3. Overview of the Employed Emulation Approach

This section provides a high level description of the e2e emulation approach adopted in the evaluation platform. It is worth noting that two different types of users are considered in the testbed, namely the UUT (a single user with real IP traffic that is truly transmitted through the whole testbed from the server PC to the client PC), and the rest of users (referred to as emulated users), whose traffic is artificially generated according to accurate traffic models and injected at specific points of the testbed in order to reproduce certain load conditions in the network. Depending on the instantaneous conditions of the network, the real IP traffic for the UUT is impaired throughout the e2e transmission and the resulting QoS degradation is evaluated in real-time.

In order to explain the packet-data flow for the UUT, let's assume as an example that the UUT has an active session in the system, is connected to UTRAN, and UTRAN is attached to the CN through IR1 (this association is configurable). The process will be explained for IP packets transmitted from the server PC to the client PC, i.e. in downstream; for upstream the process is analogous. In downstream, IP data packets generated by the applications server enter the CN through the ER. Since the UUT is connected through UTRAN, and UTRAN is attached to IR1, the IP packet will be routed towards IR1. In this process, each CR PC will act as a conventional router, i.e. upon the reception of an IP packet a routing table is queried and the packet is forwarded to the next hop. The actual path followed by packets, i.e. through CR2 or through CR3-CR4, depends on several aspects such as the instantaneous conditions and the operator QoS and resource management policies (e.g., high priority packets may follow a shorter path through CR2, or may follow the path with the effective shortest delay). The unbalanced fish topology of the CN makes traffic engineering possible. The packet from IR1 is then captured by the TS and stored in a buffer. Since the UUT is connected to UTRAN, the TS informs the UTRAN emulator about the packet arrival and awaits until the emulation ends. Such emulation is performed taking into account the instantaneous RAN conditions as it will be explained in more detail in Section 5. The emulation result is then informed to the TS. Based on the obtained result, the TS may discard the packet (the packet is lost) or forward it to the UE (the IP packet arrives to the client application). The user QoS perception can then be observed in the client PC. Notice that the perceived QoS depends on the rate of lost packets and/or the delay experienced by the correctly transmitted ones, which in turn depends on the whole network design and configuration. Therefore, this approach enables an accurate e2e evaluation of real user applications over a realistic and complete all-IP heterogeneous network with advanced RRM/CRRM algorithms and e2e QoS management policies.

An additional aspect affecting the UUT perceived QoS is the instantaneous load conditions of the network. To reproduce the desired load conditions, the existence of a number of emulated users is assumed. The traffic of emulated users is not real IP-traffic as in the case of the UUT. Instead, it is artificially generated according to accurate traffic models implemented in RAN emulators. This traffic is internally used to reproduce the desired load conditions under which the UUT traffic is transmitted in the RAN emulators. In order to reproduce the same load conditions in the CN part, the RAN emulators periodically inform the TG about the instantaneous traffic amount being processed. The TG then uses this information to generate the same traffic load in the CN by injecting packets through the edge routers. For downstream emulated users, IP packets are injected in the ER and captured in the IRs. Analogously, upstream emulated traffic is injected by the TG in the IRs and captured in the ER. This approach has been proven to be appropriate in order to reproduce the desired load conditions under which the performance of the UUT traffic can be analyzed.

The previous paragraphs have provided a general overview of the AROMA testbed. The rest of this paper is devoted to provide a more detailed description of the complete evaluation platform. To this end, three main parts can be distinguished: the RAN emulation part, the CN part, and the e2e management framework. In the RAN side, the testbed implements a comprehensive real-time UTRAN/GERAN/WLAN stand-alone emulation platform that also includes the effects of the IP-transport layer due to the all-IP approach considered. The CN side comprises a realistic implementation based on a DiffServ/MPLS domain with the required QoS and mobility management functionalities. The e2e management framework considers all the relevant elements related to the e2e QoS and mobility management in the different sections: the WQB in the RAN, the BB in the CN, and the Master PDP as a hierarchical management element. It also includes the related negotiation procedures and the alignment of the QoS and mobility criteria and parameters used in the different domains. The following sections provide a detailed picture of these parts.

5. RADIO ACCESS NETWORK

5.1. General Description

The heterogeneous wireless access network implemented in the AROMA testbed is composed of three different RANs, namely UTRAN, GERAN and WLAN. The three Radio Access Network Emulators (RANEs) comprise multi-user and multi-cellular mobile environments, with physical layer emulation, standard-compliant protocol implementations, and comprehensive system-level scenarios that account for specific cell layouts, base station deployments, sectorized and omni-directional antennas with beam patterns, transmitter and receiver configurations, and large number of mobile terminals with accurate mobility and traffic models.

The different functions performed at each level of the protocol stack have been accurately implemented and modeled. Physical layer emulation is addressed by means of statistics obtained from extensive off-line link level simulations in order to reduce computational requirements while preserving a realistic behavior. The functionalities related to higher layers of the protocol stack are implemented in detail according to the specifications developed by standardization bodies. This emulation approach has been proven to be able to guarantee an accurate evaluation of the system performance under real-time constraints.

The inputs to the RANEs are essentially the scenario to be evaluated, characterized basically by the number and location of the GERAN base stations, UTRAN nodes-B and WLAN access points, the number of users per service as well as their QoS requirements, and also the specific values for the parameters of the RRM/CRRM algorithms to be evaluated. The testbed allows for the simultaneous execution of several algorithms in parallel. On the other hand, the implemented RANEs provide an exhaustive set of performance statistics that can be monitored in real-time and saved into log files. Log files allow the obtained results to be further post-processed and analyzed in more detail off-line. The RANEs operate with a time resolution of 10 ms, which has been proven to be a good trade-off between computational complexity (constrained by the real-time operation requirement) and accuracy of results.

5.2. RANEs Functional Architecture

From a functional point of view, the procedures considered in the RANEs are reflected in Figure 3. A detailed description of the indicated functionalities and their associated models is provided in the following sections.

5.2.1. Network Deployment

The *network deployment* module allows the introduction of the parameters that define the scenario to be evaluated, including service area dimensions, specific cell layouts, number of base stations and base station locations, omni-directional or sectorized antennas and their radiation patterns, and transmitter/receiver configurations among others. These parameters can be configured separately for each one of the considered RANs. Their values can be configured based on the results obtained with the aid of a radio network planning tool or can be manually selected in order to reflect the configuration of a real deployment.

5.2.2. Mobility Model

The *mobility model* module computes the trajectories of mobile terminals. Users move along a single rectangular service area the size of which is configurable (a default 8 km × 4 km configuration is employed). The UUT's movement during the emulation can be driven by the implemented mobility model or can be configured manually so that the UUT follows a predefined set of specific coordinates. The use of deterministic trajectories is valuable when trying to analyze the behavior of specific mobile terminals under test. In this case the UUT can be configured to follow the predefined coordinates one way, back and forth, or periodically. In the case of emulated users, the only allowed possibility is to employ the implemented mobility model.

During the initialization phase of the RANEs, users are uniformly scattered in the service area. This is achieved by drawing the abscissa/ordinate for each user from a uniform distribution between zero and 8 km/4 km respectively. The initial direction of the movement for each user is obtained from a random variable uniformly distributed in the interval [0, 2π).

During the emulation, users' movement is driven by a macro-cellular pseudo-random mobility model with semi-directed trajectories, which is an evolved version of the model described in [44]. Users' movement is modeled as a set of random steps with a fixed length (usually the de-correlation distance described in Section 5.2.3). Thus, the time required to cover this distance depends on the user speed, which can be configured to 0 km/h (static indoor users), 3 km/h (outdoor pedestrian users), 50 km/h (low-speed vehicular users) and 120 km/h (high-speed vehicular users). At each position update the direction may remain unchanged with a probability of 0.8 or may change with a probability of 0.2. In the latter case, the new direction is computed by adding a random angle to the previous direction. The random angle is selected from a uniform distribution between [–40, +40] degrees. A wrap around technique is applied, meaning that when a mobile reaches the boundary of the simulation area it reappears on the opposite side. As a result it always stays within the considered service area.

5.2.3. Propagation Models

The *propagation model* module computes the propagation loss experienced between different locations in the service area and the deployed base stations. Such models have been implemented in order to compute the received signal power for each user based on the transmitted power. The propagation loss is computed as the sum of two terms: the path loss and the shadowing loss. The path loss model provides an average measure of the signal attenuation over a given distance. However, for the same distance between transmitter and receiver, different values of instantaneous loss can be obtained due to different surrounding environments. This effect is included by means of the shadowing, which adds additional signal attenuation due to obstacles in the path between transmitter and receiver.

The path loss is modeled as described in [44]. Outdoor antennas have been considered in all cases, including WLAN hotspots. This yields the path loss expression

$$L(\mathrm{dB}) = A + B\cdot\log_{10}[d(\mathrm{km})] + C$$

where L is the path loss in decibels, d is the distance between transmitter and receiver in kilometers, A = 120.9 for GERAN (900 MHz), A = 128.1 for UTRAN (2000 MHz) and A = 129.8 for WLAN (2400 MHz), and B = 37.6 (assuming an approximated antenna height of 15 meters). For outdoor users C = 0, which yields the path loss model for vehicular test environment described in [44]. For indoor users an additional loss C ranging from 17 to 20 dB has been considered depending on the considered RAN, which approximately yields the path loss model for outdoor to indoor described in [44].

Experimental measurements have shown that the shadowing loss can be modeled as a random process with a normal distribution of mean 0 dB and standard deviation between 4 and 12 dB depending on the propagation environment. A shadowing standard deviation of 10 dB has been considered in this work. The shadowing is a spatially correlated process, meaning that the shadowing losses experienced by a mobile terminal at two nearby positions are correlated. This spatial correlation has been modeled as detailed in [45] with a decorrelation distance of 20 m.

The usual approach when simulating a mobile communication system is to compute the propagation loss every time it is required during a simulation. However, due to the stringent real-time requirements of the RANEs, a different approach has been adopted in order to reduce the computational load. A propagation matrix (specific for each emulated scenario) is computed during the initialization phase of the testbed. This matrix contains the propagation losses experienced between each one of the base stations included in the emulation scenario and a grid of discrete geographical locations. Every time a user’s position is updated by the mobility model, the new location is approximated to the nearest point of the propagation matrix, and the new propagation conditions are derived from the matrix. The distance between consecutive points of the propagation matrix is equal to the decorrelation distance of the shadowing model. Notice that all the possible propagation conditions are computed during the initialization of the RANEs and read during the emulation, instead of computing them in real-time. This approach reduces the real-time computational requirements and therefore results more appropriate for real-time emulation.

5.2.4. Horizontal and Vertical Handover Decisions

Based on the received signal power obtained from the propagation models, the *horizontal handover* and *vertical handover* modules may trigger a handover. Handover decisions are based on the received signal power for the common broadcast channels, i.e. the Common PIlot CHannel (CPICH) for UTRAN, the Broadcast CHannel (BCH) for GERAN, and the periodic beacon signals for WLAN. A new cell is considered as reachable by a given mobile terminal when the received signal power for its common broadcast channels exceeds a predefined threshold. If the received signal power from the new cell is greater than the old cell's received signal power plus a hysteresis margin, a handover is triggered, which may be a horizontal handover (between cells of the same RAN) or vertical handover (between cells of different RANs). The processes for horizontal handovers of GERAN [46] and UTRAN [47], as well as vertical handovers between them or with other technologies [48] are defined in their corresponding technical standards and specifications.

5.2.5. Traffic Models

The use of accurate and realistic traffic models is of paramount importance when evaluating RRM/CRRM procedures. In effect, traffic generation involves deciding the instants when users start and finish sessions as well as the instants when data packets are generated and buffered. The start of a new session will trigger an admission control procedure to check if the user can be accepted depending on the system status. Similarly, the number of packets remaining in the buffers will be used by the Medium Access Control (MAC) and scheduling algorithms to select the appropriate radio transmission parameters. Therefore, accurate and realistic traffic models are required when studying RRM/CRRM solutions.

The testbed implements voice, video streaming and web browsing traffic models as practical cases of conversational, streaming and interactive services, respectively. Traffic generation is driven at two different levels. The first level determines whether the user has an active session of a given service. When the user's session is active, the second level determines the time instants when data packets are generated and their size.

The session level is modeled in the same way for all the implemented services. The beginnings of the first session for all emulated users are scheduled during the initialization phase of the testbed and are uniformly distributed in time during the first instants of the emulation (usually the first 100 seconds). This approach allows increasing the number of active users linearly and avoids an excessively high number of simultaneous active users at the beginning of the simulation, which is constrained by the computational capabilities of the RANEs. Mobile wireless subscribers are usually considered to have independent behavior one from each other, which results in exponentially distributed session inter-arrival times. Session durations can also be modeled by an exponential distribution. It is widely agreed that a negative exponential distribution is a good approximation for such stochastic processes distributions. Therefore during normal operation of the RANEs the time period between the start of two consecutive sessions, i.e. the session inter-arrival time, and the session duration are both modeled by means of a negative exponential distribution. Consequently, traffic at session level is modeled by means of a Poisson session arrival process and an exponentially distributed session duration random variable with a predefined mean. For instance, typical values employed for a voice user are 1 call/hour with an average duration of 120 seconds/call. For other services different values may be selected, depending on the service's characteristics and user profiles to be analyzed or the specific load conditions desired in the RANEs. It is worth noting that the session generation process may be interrupted for different reasons, e.g. when a dropping condition holds as a consequence of insufficient resources or non-fulfillment of QoS requirements.

When the user's session is active, the second traffic modeling level drives data generation according to service-specific traffic models. The voice traffic is implemented as a sequence of consecutive talk spurts (active) and silence (inactive) periods within each call as shown in Figure 4. The duration of the active and inactive periods was studied in [49], revealing a negative exponential distribution for the duration of both active and inactive periods. These distributions are characterized by the mean duration for the active periods $\overline{T}_{active}$ = 1.35 seconds, and the mean duration for the inactive periods $\overline{T}_{inactive}$ = 1.70 seconds. The activity factor of the source is defined as the proportion of the time that the source is active. For negative exponential distributions, the activity factor can be extracted as $\overline{T}_{active} / (\overline{T}_{active} + \overline{T}_{inactive})$. During active periods, data packets are generated at a constant rate (e.g., for a 12.2 kbps voice service a data packet of 122 bits is generated every 10 ms during active periods).

The video streaming traffic is implemented similarly to the voice service with an activity factor of 100%, i.e. the source is active during all the session lifetime and packets are generated at a constant bit-rate (assuming a constant bit-rate streaming service).

The web browsing traffic model implemented [44] is depicted in Figure 5. A web browsing session is composed of several web page transmissions, referred to as *packet calls*. A packet call is initiated by the submission of an URL request by the user. The packet call consists in the transmission of several datagrams containing the web page's HTML objects. When all the objects are downloaded the user then spends some time reading the web page before performing another URL request, which in turn initiates another packet call. This model is characterized by the following parameters: number of packet calls per session N_{pc}, reading time between packet calls T_{pc}, number of datagrams per packet call N_d, inter-arrival time between datagrams within a packet call T_d, and datagram size S_d. The datagram size S_d is modeled by a Pareto distribution with cut-off:

$$f_x(x) = \begin{cases} \dfrac{\alpha k^{\alpha}}{x^{\alpha+1}}, & k \le x < m \\ \beta, & x = m \end{cases}$$

where $\alpha = 1.1$, $k = 81.5$ bytes, $m = 66666$ bytes is the maximum allowed datagram size, and $\beta = (k/m)^{\alpha}$, $\alpha > 1$, is the probability that $x > m$. The parameters N_{pc} and N_d have been modeled as geometrically distributed random variables with mean values $\mu_{Npc} = 5$ and $\mu_{Nd} = 25$ respectively, while T_{pc} and T_d have be modeled as exponentially distributed random variables with mean values $\mu_{Tpc} = 20$ seconds and μ_{Td}. The inter-arrival time μ_{Td} is adjusted in order to obtain different average bit-rates at the source level taking into account the average datagram size μ_{Sd}, which is computed as $\mu_{Sd} = [\alpha k - m(k/m)^{\alpha}]/(\alpha - 1)$.

Before concluding, it is worth remarking that the described traffic models are employed to generate traffic for emulated users. The UUT traffic is real IP traffic generated at the applications PC.

5.2.6. Radio Resource Management Functions

The CRRM module is in charge of the set of functions that are devoted to ensure an efficient use of the available radio resources in the heterogeneous scenario by means of a proper coordination between the different RANs. The CRRM module is responsible for coordinating the execution of vertical handovers between RANs (RAN selection) as well as the individual RRM strategies for each RAN (local RRM configurations). The local RRM modules for UTRAN, GERAN and WLAN are in charge of the set of functions devoted to manage the resources of each individual RAN, i.e. admission control, congestion control, power control, link adaptation, scheduling, resource allocation, and so on. A wide set of algorithms for each one of them is implemented in the testbed, including by default the classical and most widely employed RRM/CRRM algorithms for each radio technology. The description of such techniques and associated algorithms is out of the scope of this work due to their complexity, technology specificity, and number of implemented solutions. The interested reader can however find the most widely employed algorithms for these RRM/CRRM techniques in the existing rich literature on this field [1][2][3][4][5][10]. It is worth noting that the RRM/CRRM algorithms implemented in the testbed by default are not an inherent part of the testbed itself. Therefore, any novel RRM/CRRM technique could be included in the testbed in order to evaluate its performance in a realistic environment.

Although only a single UUT is running real applications on the testbed, it is worth noting that RRM/CRRM algorithms are applied indistinctly over all the traffic processed by the RANEs, including the traffic of both the UUT and the rest of emulated users. Therefore, the UUT behaves as any other user in the system where processing of the data differs along time depending on the current RAN status, i.e. load conditions, interference, etc.

5.2.7. Signal and Interference Computation

The *signal and interference computation* module computes, for all the users, the transmitted power and the experienced interference in order to obtain the corresponding signal to interference ratio at the receiver. The signal level is computed in a similar way for all the RANs. However, the interference level computation depends on the considered RAN.

For Frequency/Time Division Multiple Access (FDMA/TDMA) systems such as GERAN, interference proceeds from co-channel cells separated from the interfered cell by a given reuse distance. Therefore, the GERAN channel quality can be expressed by means of the Carrier-to-Interference Ratio (CIR) as follows:

$$CIR_{GERAN} = \frac{\dfrac{P_i}{L_P^{ii}}}{\sum_{j \in \Omega} \dfrac{P_j}{L_P^{ij}} + N_0 \cdot W}$$

where P_i is the transmission power of the desired signal in the user's cell (cell *i*), L_P^{ii} is the propagation loss between the base station and the user in cell *i*, Ω is the set of active transmitters in co-channel interfering cells, P_j is the transmission power of the co-channel interfering users, L_P^{ij} is the propagation loss between active transmitting interferers in cells *j* and the interfered user in cell *i*, and $N_0 \cdot W$ represents the thermal noise at the receiver in cell *i*, with N_0 being the thermal noise spectral density (– 174 dBm/Hz) and *W* the bandwidth of the transmission channel (200 kHz for GERAN).

For Code Division Multiple Access (CDMA) systems such as UTRAN, an additional intra-cell interference component exists as a consequence of multipath propagation, which decreases the orthogonality between the channelization codes of the cell. Intra-cell interference on a CDMA system is modeled by an orthogonality factor α [50]. In absence of multi-path fading, the codes are perfectly orthogonal and $\alpha = 1$. A value $\alpha \approx 0.5$ means that two different samples of the same signal are received with similar strength. In the worst case $\alpha = 0$, meaning that orthogonality is entirely destroyed. Typical values are between 0.4 and 0.9 [51]. Therefore, the UTRAN channel quality can be expressed as:

$$CIR_{UTRAN} = \frac{\frac{P_i}{L_P^{ii}}}{\sum_{j\in\Omega} \frac{P_{T_j}}{L_P^{ij}} + \frac{(P_{T_i} - P_i)(1-\alpha)}{L_P^{ii}} + N_0 \cdot W}$$

where P_{T_i} and P_{T_j} represent the total transmitted power in the considered cell i and in the interfering cells j, respectively. In the case of UTRAN, W = 5 MHz.

In the case of WLAN, it is assumed that the distance between hotspots results in no interference between WLAN users. As a result, the channel quality is represented by the Signal-to-Noise Ratio (SNR):

$$SNR_{WLAN} = \frac{\frac{P_i}{L_P^{ii}}}{N_0 \cdot W}$$

The signal and interference levels experienced by each user are computed after every transmission in order to determine the experienced channel quality and, based on this parameter, to decide whether the transmitted information is correctly received. To this end, a set of results obtained from off-line link level simulations are employed, as explained in Section 5.2.8.

5.2.8. Statistics from Link Level Simulations

The simulation of mobile and wireless communications systems is usually split into two different levels, namely the link level and the system level. The link level focuses on the physical layer behavior of the channel used by a mobile user to communicate with its corresponding base station, either in the uplink or in the downlink. Link level simulations are performed with a time resolution in the order of bits or channel symbols and are aimed at characterizing the channel performance in terms of transmission error rates. On the other hand, system level simulations focus on the behavior of RRM/CRRM algorithms in a multi-cell, multi-user and multi-service scenario, as it is the case of the RANEs implemented in the AROMA testbed. By contrast, system level simulations are performed with a much higher time resolution, in the order of time-slots or radio frames. To handle this complex scenario in moderate simulation times both levels are simulated independently and the system level simulator makes use of the off-line results obtained by means of the link level simulator. The outputs obtained from link level tools are mainly concerned with the BLock Error Rate (BLER) or Frame Error Rate (FER) as a function of the channel quality for different transmission conditions (modulation and coding schemes, user speeds or propagation environments). After each radio transmission in the system level simulator, the experienced channel quality is computed as indicated in section 5.2.7. The error rate X_0 corresponding to the obtained experienced channel quality is then determined with the aid of the link level results. Then, a random number X is drawn from a uniform distribution between zero and one. If $X > X_0$ the transmitted data block is assumed to be successfully received. On the other hand, if $X \leq X_0$ the transmitted data block is then assumed to be received in error.

To minimize the impact of using off-line link level simulations on the reliability of the obtained results, and to accurately account for instantaneous channel quality variations, the real-time wireless transmission is emulated at the packet/slot level. This means that the decisions on success/failure of transmissions are not based on average channel quality values (e.g. average CIR experienced during a packet transmission), but in the channel quality distribution along the various slots required to transmit the packet, which enables to reasonably reflect the impact of the instantaneous channel quality variations on the transmission results.

5.2.9. IP-RAN Emulation Model

In all-IP networks, IP transport is employed not only in the CN part, but also in the RAN part. Existing legacy interfaces are kept but they are supported over an IP-based packet-switched network. As a consequence of such approach, a data block can be lost not only in the radio interface because of unfavorable radio conditions but also due to transport network losses or excessive delays. Therefore, a model for representing the effects and impairments of the IP-based transport in the RAN becomes necessary. The envisaged IP-RAN emulation model takes into account losses in the transport network, obtained from non-real-time simulations, as shown in Figure 6. In these off-line simulations, a data block is discarded in the IP transport network if the experienced delay is higher than a predefined threshold. The loss statistics depend on the IP-RAN topology chosen, the traffic and user mobility patterns, the dimensioning of the network as well as the QoS and IP mobility architecture chosen (over-provisioning, pure

DiffServ or QoS routing). These loss statistics obtained from off-line simulations are used to determine, for different scenarios, the probability that an IP packet is lost in the IP-RAN transport network, which is used to decide in real-time whether a packet is discarded due to IP transport impairments. For more details on this model, see [52].

5.3. RANEs Execution Loop

The RANE PCs and their neighbors (CRRM and TS) have five operating states: *setup*, *init*, *run*, *pause* and *stop*. The execution flow through these states is controlled from the AGMT with the aid of the CM. During the *setup* state the AGMT verifies the reachability of all the RANE machines and establishes the communication with each one of them. During the *init* state the processes executed in each PC read configuration parameters and initialize resources (e.g., memory allocation and communication flows with other modules). Once modules are initialized, they switch to the *run* state, which is the normal execution state and will be detailed later on. While the testbed is running, the execution may be paused by entering the *pause* state (processes in execution are frozen) in order to allow the captured statistics to be visually analyzed. Finally, when the testbed is stopped, all the resources in use by the processes are released during the *stop* state.

The run state is the normal operation mode. During this state, several emulation PCs execute a set of specific operations that are periodically repeated every 10 ms. Execution loops of these machines run perfectly synchronized, i.e. at the beginning of every new 10 ms period, all the synchronized modules execute their respective operations and then remain idle until the beginning of the next 10 ms interval. Synchronization is accomplished with the aid of a special process, which runs on all the machines that must be synchronized. Notice that the time resolution provided by the 10 ms execution period may be appropriate for emulating in real-time the Transmission Time Interval (TTI) of certain technologies (e.g., 10, 20, 40 and 80 ms in UTRAN), but may be insufficient for other technologies with shorter TTIs such as e.g. HSPA (2 ms). To emulate technologies with TTIs that are *N* times shorter than the 10 ms execution period, a virtual execution loop *N* times shorter is obtained by repeating the corresponding operations and computations *N* consecutive times within the 10 ms execution period. Notice that this approach maintains the overall 10 ms execution period of the RANEs unaffected and only implies slight modifications to those technologies with shorter TTIs. This approach was proven to allow an adequate real-time emulation of shorter TTIs, with negligible effects over the real-time performance of the transmitted traffic, while leaving the real-time emulation of the rest of RANE functionalities unaffected. For more details the reader is referred to [30].

The execution loop followed during the run state of the RANEs is qualitatively illustrated in Figure 7. As it can be appreciated, the first operation executed every iteration is to update mobiles' positions within the scenario, according to the mobility model described in Section 5.2.2. Afterwards, the propagation conditions for the new positions are updated using the models presented in Section 5.2.3. Depending on the new propagation conditions, the set of reachable base stations may be updated and a handover between cells may be performed as described in Section 5.2.4. The next step is to read and process messages coming from CRRM, if any. Messages from the TS informing about the arrival of a new IP packet for the UUT are also read and processed. For the rest of emulated users, the traffic is generated according to the traffic models described in Section 5.2.5. Later on, all users with non-empty buffers are added to a list of users requesting a transmission. Before scheduling them, the received user reports are processed since this information might be required by the scheduling algorithm employed. The implemented scheduling approach is divided into two phases. In the first phase, the list of transmission requests for each base station is ordered according to the service type: conversational services are of highest priority, then streaming, and finally interactive. In the second phase, some users are then selected by the scheduler according to the specific scheduling criterion employed. The traffic load produced by the selected users is then communicated to the TG. Afterwards, the radio transmission of each selected user is emulated taking into account the cell site deployment, number of emulated users, mobility patterns, propagation impairments and RRM functions as indicated in Section 5.2.6. All these factors determine the actual channel quality experienced during a radio transmission, which is computed as described in Section 5.2.7 and used to decide whether the transmitted information is received in error making use of link level results as discussed in Section 5.2.8. Whenever an IP packet is completely transmitted without errors, a forward message is sent to the TS.

Similarly, if either some parts of the IP packet are lost or the IP-RAN model described in Section 5.2.9 indicates that an IP packet is discarded, then a discard message is sent to the TS. Finally, user reports are transmitted in order to provide useful information for the next iteration of the execution loop.

5.4. Interaction between the RANEs and the Rest of the Testbed

The interaction between the RANEs and the rest of the AROMA testbed is accomplished through two different planes: a control plane and a data plane. The control plane supports all the functionalities needed for exchanging control messages between the RANEs and other modules, concretely the CRRM and TG machines. To this end, the CM abstraction layer provides the logical concept of *flow*. During the initialization phase of the testbed, the modules create a flow with all the modules to/from which control messages need to be sent/received. Thereafter, modules can write control packets in the flow addressed to the destination module. This message exchanging process is managed by the CM in a completely transparent way. For more details see [43].

As shown in Figure 1, the RANEs manage two control interfaces, one with the CRRM PC and other with the TG PC. The control interface with TG is used by the RANEs to communicate periodically to the TG the instantaneously experienced traffic load for each service. This information is used by the TG to load the CN with a traffic level according to that experienced in the RANs. On the other hand, the control interface with the CRRM is mainly used for session management. A session activation message is sent from CRRM to one of the RANEs whenever a new session is established within the corresponding RAN (see Figure 8). This event may occur for the UUT when a new session is activated through the QoSClient graphical interface and for emulated users according to the traffic generation model implemented for each service. For all users, a session activation message for a given RANE may also result from a vertical handover from any other RAN. This message carries information about the user, the required service type and QoS requirements, the base station the user should be attached to and the geographical position of the user in the scenario. Similarly, a session modification message may be sent from CRRM to a RANE whenever the QoS requirements change (due to an e2e QoS renegotiation event or because the UUT requests new QoS parameters) or the mobile position has to be updated (the periodicity of such updating will depend on the mobile speed). A session deactivation may be initiated by both the CRRM and the RANEs. Deactivation from CRRM, illustrated in Figure 9, occurs for the UUT when the session is deactivated through the QoSClient graphical interface. For emulated users, the traffic generation models implemented in the RANEs decide the end of each session and advertise the CRRM as shown in Figure 10. Finally, the RANEs may decide to drop a specific established session for any user, e.g. if QoS requirements are not satisfied. The message exchanging in this case is illustrated in Figure 11.

The data plane comprises all the functionalities needed to support the transmission of real IP packets for the UUT through the testbed. As shown in Figure 1, a data interface between the RANEs and the TS module is defined. The interaction between these two modules is qualitatively illustrated in Figure 12. In the downstream direction real IP packets of the UUT coming from the CN are captured by the TS and are stored in a data buffer. Some descriptive parameters regarding the packet (e.g., a packet identifier, the packet size, the service type, or the QoS requirements, among some others) are sent to the corresponding RANE by making use of the communication interface provided by the CM. The RANE maintains a data structure emulating buffers of all the users (UUT and emulated users). This data structure is updated upon the arrival of a new real IP packet (indicated by the TS message, only for the UUT) or whenever a new data packet is generated by the traffic generation models implemented in the testbed (for emulated users). The radio transmission of each packet is emulated on a radio-block basis taking into account several system-level aspects such as the cell site deployment, number of emulated users, mobility patterns, propagation impairments, and so on. After the radio emulation, the result of an IP packet transmission is communicated back to the TS through the interface provided by the CM. The message sent to the TS indicates a packet identifier and the result of the emulation, i.e. correct or incorrect transmission. Then, the TS forwards the packet to the UE or discards the packet, depending on the transmission result obtained in the RANE emulation. The described procedure also applies for IP packets in the upstream direction. This procedure is completely managed in real-time. Notice that this emulation approach is able not only to reflect the loss of packets incorrectly transmitted (some packets arriving to the TS are not forwarded to their destination) but also the real-time delay experienced by packets correctly transmitted (the TS does not forward a packet until the RANE indicates it).

6. CORE NETWORK

6.1. DiffServ-MPLS Architecture

The CN implemented in the AROMA testbed is not emulated, as it is the case of the RAN. Instead, the CN part comprises a realistic implementation based on a DiffServ [36] domain, where several PCs work as true routers using the communication protocol stack of the Linux OS, enhanced with MPLS [37] forwarding support.

DiffServ and MPLS may be understood as complementary methods. The MPLS mechanism is developed to enable multiple path usage for traffic forwarding, which is achieved by inserting/removing labels at the IRs/ERs of a MPLS domain – i.e. Label Edge Routers (LERs) in MPLS terminology. Each label defines a path for a traffic flow, referred to as Label Switched Path (LSP). Inside the MPLS domain packets are forwarded through LSPs using these labels. LSPs are unidirectional, meaning that two LSPs, one in upstream and another in downstream, need to be established for bidirectional communication between client and server PCs. While MPLS enables traffic engineering, resource reservation, fault tolerance and optimization of transmission resources in the CN, it does not define a QoS architecture itself. DiffServ does define a scalable QoS architecture with multiple classes of services [36], namely Expedited Forwarding (EF) [53], several grades of Assured Forwarding (AF) [54] for different QoS requirements in terms of throughput, delay, loss and jitter, and Best Effort (BE). The 3GPP conversational, streaming and interactive classes are usually mapped to EF, higher AF, and lower AF or BE classes, respectively. IP packets entering a DiffServ domain are assigned a class of service at the domain edge, which is written in the Differentiated Services Code Point (DSCP) value in the IP header. This value defines the per-hop behavior for that class along the

CN and therefore the per-hop treatment of IP packets in terms of scheduling and queue management at each CR.

Currently there exist two main solutions for MPLS support of DiffServ [36]: E-LSP and L-LSP. In the former approach, all the classes of one flow are forwarded through the same labels. In the latter solution, labels determine different paths for different types of classes independently by combining MPLS labels and DSCPs. The L-LSP approach was selected for the AROMA testbed since it provides the required functionalities with better control over the MPLS suite software.

In this scenario, upon arrival of an IP packet the edge routers, i.e. Label Edge Routers (LERs) in MPLS terminology, look up the DSCP and IP addresses present in the IP header and determine the MPLS Forward Equivalency Class (FEC) the packet belongs to, thus deriving the corresponding LSP that the IP packet will follow along the CN. At each CR, the treatment received by each IP packet will depend on the marked DiffServ class of service.

Resources in the DiffServ domain with MPLS forwarding are controlled by the BB. In this sense, the BB is in charge of controlling LSP creation and release when certain events occur (e.g., new session establishment, session conclusion, session dropping or mobility issue) and mapping incoming traffic flows to existing LSPs. To this end, the BB needs an updated knowledge of the instantaneous CN resource usage and topology, i.e. the logical connections between routers in the CN, which in the case of AROMA testbed is the knowledge of all the previously established LSPs and their availability to receive new sessions or switch sessions between IRs. In order to collect such updated information in real-time and to properly configure the CRs, a software component referred to as *BB agent* is implemented in each CR. BB agents play the role of Policy Enforcement Points (PEP), configuring MPLS and forwarding parameters in CRs according to the instructions received from the BB and establishing appropriate filters in the ERs in order to properly classify and mark incoming IP packets. The BB manages the CN LSPs by sending appropriate configuration commands to the BB agents running in each router.

It is worth noting that conventional IP routing protocols running on the CN would create forwarding data bases that would most likely direct packets along the shortest path.

Hence, conventional IP routing would set up a route between edge routers following the path through CR1 and CR2, since this is the shortest path. Moreover, IP packets would be transmitted without QoS guarantees since all packets would be processed by CRs in the same way regardless of the service's QoS requirements. On the other hand, with the implemented DiffServ/MPLS architecture, different LSP tunnels can be set up between edge routers, as shown in Figure 13. In this case, the slightly unbalanced fish topology adopted in the CN enables traffic engineering applications since different types of traffic may follow different paths in the CN. For example, voice traffic may follow the shortest path (CR1 – CR2) while data traffic may be forwarded through the longest path (CR1 – CR3 – CR4). This path differentiation is important due to the different QoS levels that may be experienced through different paths. Moreover, the DiffServ technology is another aspect determining the QoS level experienced by IP packets and, hence, by applications run by the UUT.

6.2. Coordinated Traffic Generation

As it was discussed in section 4.3, there is a single reference user in the AROMA testbed (the UUT) running real commercial IP multimedia applications. The traffic generated by such applications is processed by the RANEs and forwarded through the CN. The rest of users competing with the UUT for system resources are emulated by means of traffic modeling. Traffic models are implemented in the RANEs (see section 5.2.5). Since the CN is based on a real implementation, traffic of emulated users needs to be generated and injected into the CN. The entity in charge of such function is the TG machine. The TG periodically collects the information sent by the RANEs and injects IP packets so that the traffic conditions experienced in the CN match those of the RANEs.

Different approaches may be considered to obtain relevant data from the RANEs. A first option would be to identify and implement in the TG a limited set of tunable traffic aggregation models. The RANEs would measure those parameters needed to characterize such aggregation models and they would provide the value of these parameters to the TG machine. The aggregation models running in the TG would then be configured with these values in order to inject to the CN the appropriate traffic load. Another option would be to implement in the

TG the same traffic models implemented in the RANEs. In this case the RANEs would measure the number of active users per service in each RANE and they would send this information to the TG periodically. The TG would then handle an equivalent number of traffic sources that would be in charge of generating live traffic to be injected to the CN. A simpler approach has been adopted in the AROMA testbed, the principle of which is to periodically measure and send to the TG the instantaneous amount of processed data bytes per service in each RANE. This information is used by the TG to directly inject the equivalent traffic load.

Injected IP packets are generated by a process that creates real data packets and inserts them into the CN according to the aggregated traffic indicated by the RANEs. In order to ease the control of traffic differentiation per service class and the control of the attachment point (IR) of each RAN, separate flows are generated for different services in each RAN (see Figure 14). The TG thus controls up to 18 real traffic flows. For downstream flows, IP packets are injected in ER and removed in the IR they are forwarded to. Analogously, IP packets for upstream flows are injected in the appropriate IR and removed in the ER. IP packet sizes are predefined and fixed for a certain class; the appropriate traffic load is controlled by generating the proper number of IP packets. IP packet sizes as well as RANEs report periodicity are tuned in order to obtain the desired performance.

7. END-TO-EDGE FRAMEWORK

The two previous sections have been devoted to the description of the RAN part (Section 5) and CN part (Section 6) of the testbed. There exists a third important component, the e2e framework, which comprises all the procedures and functionalities that jointly and coordinately involve the RAN and CN parts. Two main groups of functionalities are addressed from an e2e point of view, which are related to QoS and mobility issues. These functionalities are needed to provide an environment with QoS support to mobile services. Both aspects will be described in sections 7.1 and 7.2 respectively. Although these two features are presented in separate sections, there actually exists a tight relation and interaction between them.

7.1. QoS Management

The implemented e2e QoS management framework is based on Policy-Based Networking (PBN), in which network wide policies to be enforced in each domain are obtained as a result of a negotiation process among the involved domains. The PNB approach implemented in the testbed is that of an all-IP B3G heterogeneous network, composed mainly of two different segments: the RAN domain and the CN domain. As illustrated in Figure 15, the policy-based e2e QoS management framework is mainly supported by the WQB and BB entities in the RAN and CN parts, respectively, with the Master PDP entity coordinating both domains and making the final e2e decisions. Additionally, other entities such as the QoS Client (the QoS negotiation application of the UUT) and the CRRM module (in charge of coordinately managing the radio resources of different RANs) are also involved in the e2e QoS procedures. The e2e QoS support is enabled by the proper interaction among these entities. Interaction is envisaged in terms of QoS negotiation. Therefore, QoS requirements for the whole B3G network domain are provisioned during the session lifetime accordingly in the RAN and CN parts as a result of this negotiation.

In the heterogeneous RAN, the WQB is in charge of the QoS management functions, while the CRRM module is responsible for the coordinated and efficient management of the pool of radio resources provided by each one of the individual RANs. More specifically, the WQB/CRRM performs three main functions. First, since each RAN may have specific QoS mechanisms, the WQB is responsible for monitoring and configuring the QoS mechanisms of each RAN in order to achieve the appropriate QoS provisioning. Thus, the WQB configures QoS mechanisms in RAN elements according to a set of common policies. Similarly, CRRM functions may also be configured from the WQB if required. Second, CRRM functions play a crucial role in the RAN and thus are jointly managed by the WQB/CRRM modules in order to guarantee the required QoS level during admission control and initial RAN selection procedures as well as preserve the provided QoS level during handover executions or other reconfiguration events. Finally, the WQB/CRRM takes local decisions on the QoS management in the RAN part whenever required (e.g., during admission control and/or handover procedures) according to radio resource usage, network topology and traffic distribution constraints, and then dynamically negotiates QoS agreements with the CN part.

In the CN, the BB is in charge of coordinately managing the resources and the QoS mechanisms associated to the CN. The BB is the main architecture element of the control plane of the DiffServ model for supporting e2e QoS in IP-based networks. The internal structure of the implemented BB is depicted in Figure 16. As it can be appreciated, several databases are maintained in order to provide up-to-date information about the CN to several internal modules when required. The *Policies* database is a repository containing the set of policies considered by the network operator. The *Reservations* database is an updated list of resource allocations in the CN, which is updated every time a new user session is accepted or after resource re-allocations caused by mobility events. The *Network status* database maintains up-to-date information of the actual network status and is updated by the *Measurement analyst module*, which periodically polls the CN elements in order to obtain the desired information. The *Topology* database is updated by the *Routing protocol analyzer* module, which receives and processes routing messages from the CN in order to infer the current CN topology. The *Resource request attendant* module parses and understands the resource requests performed by the WQB entity, and triggers the admission control process in the *Admission control* module. Similarly, the *Mobility attendant module* receives mobility requests, triggers the admission control process to verify if the user can move sessions to another IR, and then manages mobility events according to the current network status and topology. To this end, LSPs are created on demand in order to accommodate the requested traffic. The *Admission control* module implements admission control algorithms, responsible to accept or reject resource requests, according to a certain criterion, based on the information provided by the BB databases. Admission control is the mechanism used to evaluate whether requested resources are available in the CN or, more precisely, if the routers in the traffic path have enough resources available to support new traffic flows with the requested QoS levels. The admission control criterion can be based not only on bandwidth constraints but also on the user profile or on QoS parameters such as as jitter, delay or packet loss rate. The admission control procedure is triggered whenever a resource request is received from the WQB for a new user session entering the network (*Resource request attendant module*), or after a mobility event for an existing one (*Mobility attendant module*). After the reception of such requests, the *Admission control* module computes the path to be followed by the packets between IR and ER, associates the traffic type to a specific LSP, and checks whether the traffic characteristics and QoS requirements can be attended in the designated path. Based on this verification, the resource request is either accepted o rejected in the CN.

While the WQB/CRRM and BB entities make QoS and resource management decisions for the RAN and CN domains respectively, the final e2e decisions are made by the Master PDP entity based on the local information provided by the WQB/CRRM and BB entities. The Master PDP thus manages the negotiation of the QoS criteria and parameters used in the different domains as well as the QoS negotiation with external peer domains involved in the provisioning of end-to-end services (see Figure 15).

Different QoS negotiation mechanisms have been implemented in the testbed. To this end, a proper interface between the QoS interacting entities has been developed. The inter-domain QoS signaling is carried out by means of an interface that finds its roots in the COPS-SLS framework [55], which is a hierarchical client-server protocol that defines a Policy Decision Point (PDP) and a Policy Enforcement Point (PEP). The implemented e2e QoS signaling is based on a three-handshake signaling procedure between entities that enables the exchange of QoS parameters and decisions. Any negotiation between two entities is performed by exchanging three messages: REQuest (REQ), DECision (DEC) and RePorT (RPT). Any negotiation interaction between the QoS entities is initiated by a REQ message which encapsulates the session identifier, the flow attributes (source and destination IP addresses and ports), the performance attributes (including the requested QoS level in terms of throughput, packet loss, delays and DiffServ code point) and the conformance attributes (needed for the traffic shaping in the IRs). The entity receiving the REQ then replies with a DEC message indicating whether the QoS request can be supported or not. Finally, the negotiation is closed with a RPT message that originates the enforcement of the negotiated QoS if the negotiation was successful. Based on this signaling framework, several QoS mechanisms have been implemented, including session activation triggered by the UUT, session deactivation triggered either by the UUT or by the network in case of dropping conditions (e.g., loss of coverage, change of network preferences, etc.), and session modification (QoS renegotiation) triggered by the UUT (e.g., when UUT decides to request a higher or lower QoS level) or by the WQB/CRRM or the BB (e.g., when current QoS level is no longer supportable due to changes in the RAN or CN load conditions, or in order to accept a new incoming session at the expense of reduced QoS levels for some users). For a detailed description of the implemented procedures the reader is referred to [56].

7.2. Mobility Management

The mobility management functionality included in the testbed provides QoS-aware IP micro-mobility. Macro-mobility approaches such as Mobile IP [57] incur in excessive signaling between the Mobile Node (MN) and its correspondent node each time the MN changes its current point of attachment and a new care-of address has to be assigned by the correspondent node. This causes additional delays, packet losses and signaling overheads. Micro-mobility protocols were introduced to manage the IP mobility within macro-mobility domains (i.e., within the control area of the same correspondent node). Micro-mobility protocols can be classified into tunnel-based and host-based forwarding protocols [58]. Tunnel-based protocols follow a hierarchical architecture where the correspondent node, also referred to as Anchor Point (ANP), establishes tunnels (usually IP-in-IP tunnels) to the Access Routers (AR), i.e. MN attachment points. Hierarchical Mobile IP (HMIP) [59] and BRAIN Candidate Mobility Management Protocol (BCMP) [60] are examples of this kind of protocols. By contrast, in host-based forwarding protocols, each router in the path maintains a database whose information about the location of the MN is employed to forward packets to the MN. Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [61] and Cellular IP [62] are examples of these protocols.

In the testbed, a tunnel-based micro-mobility strategy with QoS extensions has been implemented. The BCMP protocol is used, but MPLS tunnels are created instead of IP-in-IP tunnels. When compared with other micro-mobility protocols, MPLS-based micro-mobility protocols show several advantages due to the MPLS technology, including simple forwarding decision based on a simple label, possibility of using constraint-based routing in order to better utilize network resources, creation of Virtual Private Networks (VPN) and network reliability. Furthermore, the MPLS technology has been widely adopted by operators in their networks.

7.2.1. Mobility Management Architecture and Procedures

Mobility management is supported in the testbed by three entities: the ANP, the AR and the MN. The ANP is located in the ER and constitutes the master mobility management entity. The ANP assigns the IP care-of address to the MN in the login phase and communicates with the BB in the event of IP handover. The BB also controls the creation, management and switching of the MPLS tunnels and thus closely interacts with the mobility management entities to know the instant when MPLS data paths need to be switched. The AR entities are installed in the IRs and are in charge of broadcasting Route Advertisement (RA) messages so that they can be identified by MNs. Finally, the MN entity corresponds to the UE machine and is the entity in charge of triggering MPLS tunnel switching procedures whenever IR switching is required (e.g. when the MN detects an IR address change in the RA message, or when a greater measured power is received from another IR).

During the login phase, the MN sends a signaling packet to the AR mobility agent located in the IRs, which is then forwarded to the ANP mobility agent located in the ER. This signaling packet is necessary in order to setup a filter at the ER for the interception of data packets addressed to the MN. Besides the setup of this filter, during the login phase a care-of-address is also provided by the ANP to the MN. In addition, the ANP notifies the BB the MN details and the IP attachment point. When the MN starts a new session, the BB computes the QoS path throughout the CN and the corresponding Diffserv QoS reservations. The source routing is then configured at the edges (ER for downstream and current IR for upstream) in order to properly encapsulate the filtered packets.

The IP handover procedure is triggered every time a layer 2 handover involving an AR change occurs. In such a case, the MN first sends a handover message to the new AR, which is forwarded to the ANP. After processing the handover message, the ANP sends an acknowledge message to both the new and old ARs. When handover takes place, the BB is informed about the details of the new AR. In case that a session is active during the handover, the QoS level provided through the new path is recalculated by the BB and some QoS renegotiation procedures may be triggered. A context transfer (including information about the DiffServ configuration for the MN's flows) is then performed from the old AR to the new AR. Once the handover and context transfer are completed, a LSP is set up towards the new AR.

7.2.2. Handover Types

Different Handover (HO) types may be executed in the testbed as the MN moves along the scenario: Horizontal Handover (HHO), Intra-IR Vertical Handover (VHO) and Inter-IR VHO.

The HHO is the classical HO mechanism of single RAN networks where an intra-RAN HO between base stations of the same RAN is performed. This event is locally managed inside the RAN. Therefore, no e2e QoS negotiation is required.

In the intra-IR VHO case, the HO is performed between base stations of different RANs attached to the same IR/AR. Its management involves the CRRM module. By default, all packets sent to the old RAN during the execution of the VHO are discarded once the VHO is completed. Nevertheless, it is possible to forward those packets to the new RAN as well. Hereafter, this latter possibility will be referred to as the *transfer policy*.

Finally, in the inter-IR VHO case, the HO is performed between base stations of different RANs attached to different IRs/ARs. In such a case, IR switching is required and thus the e2e mobility management function plays a crucial role, as explained next. In the CN side, the MN's data packets are encapsulated into MPLS tunnels from the ER/ANP to one of the IRs/ARs for downstream, and vice-versa for upstream. The TS filters data packets in the UE interface (upstream) or IR interfaces (downstream) to pass them to the appropriate RANE for emulation. Every time a VHO involving IR switching occurs, the mobility management functions are responsible for properly urging the TS to change its configuration in order to filter MN's packets from the right interface, i.e. the interface connected to the new IR.

It is important to remark at this point that inter-IR VHOs are executed in the RAN domain regardless of the IR/MPLS-tunnel switching process in the CN domain, and always after the e2e QoS negotiation. In particular, the MN first realizes that an IR/AR change has occurred during the VHO after receiving RAs from the new IR. Then, the MN sends a MPLS tunnel (LSP) change notification message to the ANP, which in turn informs the BB, thus triggering the MPLS tunnel switching to the new IR. Therefore, inter-IR VHOs result in a misalignment (at the IRs) between the paths configured in the RAN and CN domains for a certain period during the VHO execution.

It is clear that a lack of synchronization between the IR/MPLS-tunnel switching in the CN and the VHO in the RAN may lead to significant packet losses and important QoS degradations for the final user. To avoid this situation, an advanced mobility management procedure, referred to as *HO preparation*, has been implemented in the testbed. Prior to the inter-IR VHO (concretely, when the MN receives RAs from both IRs), this procedure creates an additional tunnel between IRs (inter-IR tunnel) in order to minimize packet losses during the VHO execution, as explained in section 7.2.3.

7.2.3. Handover Preparation

In order to illustrate the inter-IR VHO procedure with HO preparation, the exchanged signaling messages are detailed in Figure 17. In this example, the UUT (MN) is connected to UTRAN through IR1 before the VHO. A logical radio path or bearer is therefore established between the UUT and UTRAN. The TS establishes an interconnection path that physically connects UTRAN with IR1 in the CN, while the BB establishes the corresponding MPLS path along the CN. Let's assume that until the beginning of this example the UUT was only under UTRAN coverage and from that moment is also under WLAN coverage. Notice that if the UUT is located in an area where there is coverage from various RANs, the MN then receives RAs from all the IRs where the RANs are connected to. The inter-IR VHO procedures are executed as follows:

1. When the MN starts receiving RAs from both IRs, it realizes that a VHO may be near to happen and then a HO preparation message is sent to the current IR (i.e., IR1). This message triggers the creation of a tunnel between the involved IRs. In the testbed, the TS emulates the tunnel by simultaneously connecting the UUT to both IRs instead of

physically creating a tunnel between them. However, in the following we refer to this mechanism as the inter-IR tunnel. Then, as long as the inter-IR tunnel is active, data packets forwarded to either IR1 or IR2 are captured and sent to UUT.

2. Next, if the RAN selection procedures executed in CRRM determine that a VHO from UTRAN to WLAN is required, a VHO request is sent by the CRRM module to WQB. The WQB initiates an e2e QoS renegotiation that finishes with a new radio bearer established to the new RAN and the UUT connected to the new IR (i.e., IR2). However, at this moment the MPLS tunnel through CN has not been changed yet.
3. When the MN realizes that RAs from IR2 are received with higher power, it requests an IR/MPLS tunnel change to the ANP, which forwards the message to the BB. As a result, a new MPLS tunnel is established to the new IR, and the old MPLS tunnel is released. Notice that the RA period (the time between two consecutive RAs) is longer than the CRRM measurement period considered to perform VHOs. As a result, VHOs are executed before IR/MPLS tunnel switching. The RA period therefore highly impacts on the time interval during which the RAN and CN paths are misaligned.
4. Finally, when only RAs from the new IR are received, the MN requests the inter-IR tunnel release.

Notice that if the inter-IR tunnel had not been created, packet losses would take place during the inter-VHO execution until the BB was informed to switch the CN MPLS tunnel from the old IR to the new IR. On the other hand, the HO preparation mechanism described in this section allows harmful packet losses to be minimized.

8. CASE STUDIES

This section provides some illustrative results showing the capabilities of the developed testbed to analyze in real-time the system performance and user perceived e2e QoS/QoE. Concretely, three different case studies have been selected in order to illustrate some examples of the kind of network research studies that can be carried out with the developed platform. Although the testbed implements a complete B3G IP-based heterogeneous wireless network, it is worth highlighting that the number of simplifications with respect to a complete real system has been minimized. As a result, the presented detailed platform also constitutes a powerful tool to study specific low level features for particular components and technologies. To illustrate this point, the first case study evaluates the behavior and performance of various scheduling algorithms for the High Speed Downlink Packet Access (HSDPA) technology. The second case study is devoted to show how the testbed can be employed to assess different mobility management strategies during mobility events such as horizontal and vertical handovers. Moreover, the impact of each one of the considered mobility management strategies on the user QoE is analyzed as well. Finally, the third case study presents an example where a complete e2e system reconfiguration takes places, involving all the testbed entities in order to provide the required e2e QoS level. Further examples and case studies can be found in [63].

8.1. Case Study I: RAT Specific Performance Evaluation

A simple case study comparing the behavior of two basic scheduling policies in HSDPA is presented. In particular, the Maximum C/I (MaxC/I) and Round Robin (RR) scheduling algorithms are considered. The MaxC/I criterion allocates resources to those UEs experiencing the best instantaneous channel quality conditions. This approach achieves high cell throughput values at the expense of user throughput distribution. Users located at cell border may not be served at all while users experiencing good transmission conditions may monopolize the resources. On the other hand, with the RR criterion resources are assigned to users on a sequential and cyclic basis; knowledge of the experienced channel quality is not made use of. This policy offers a fairer throughput distribution among users at the expense of the overall system throughput. The objective of this sample study is to show how the aforementioned behavior is evaluated in real-time and analyze the impact on the end-user perception.

In this case study, the UUT periodically moves in straight line between two base stations, thus experiencing different channel quality conditions. After 25 seconds of emulation, the UUT requests a streaming session. The session is accepted and at time instant 30 s the video sequence starts reproducing. The obtained results are shown in Figure 18 and Figure 19. The graphs on the left show (from top to bottom) the abscissa coordinate of the UUT within the service area (between two base stations), the CIR measured for the pilot in decibels, and the UUT's Channel Quality Indicator (CQI) report (the maximum CQI value for HSDPA category 12 is equal to 15 - higher CQI values correspond to the same transmission parameters). As it can be observed, the measured pilot signal changes rapidly but in average it varies according to the UUT's position: the most favorable channel quality conditions are observed when the UUT is near to a base station (time instants 0, 130 and 260 sec) while the poorest channel quality is experienced at the cell border (time instants 65 and 195 sec). This behavior is also observed in the CQI report sent by the UUT, which follows the measured pilot strength.

The behavior of the two considered scheduling algorithms can be observed in the top-right side of Figure 18 and Figure 19. These graphs show the number of Transmission Time Intervals (TTIs), measured over 10TTI/20ms periods, that the UUT requests a transmission (blue line) and the number of times that these requests are accepted by the scheduler (green line) or rejected (red line). As it is expected, the MaxC/I behavior is highly dependent on the experienced channel quality: under favorable channel quality conditions all the UUT's requests are accepted, while several requests are rejected when the UUT is at the cell border (time instants around 65 and 195 sec). On the other hand, the RR policy exhibits a more homogeneous and channel-independent pattern, with an average of around 80% transmission requests accepted (20% rejected) under both favorable and unfavorable channel quality conditions. This behavior is also observed in the number of physical channel codes, i.e. High-Speed Physical Downlink Shared Channels (HS-PDSCHs), that the scheduler assigns to the UUT, which is shown in the bottom-right side of Figure 18 and Figure 19 (measured over 10 TTI/20 ms periods). At the cell border, the number of HS-PDSCHs assigned to the UUT is limited by the maximum number of HS-PDSCHs corresponding to the reported CQI value. As a result, a low number of HS-PDSCHs is assigned to the UUT in time instants around 65 and 195 sec with both MaxC/I and RR. On the other hand, under good channel quality conditions, data transmissions can be reliably performed using the maximum number of simultaneous codes allowed by the UE's HSDPA category (5 codes for category 12). In this case, the main aspect limiting the number of HS-PDSCHs assigned to the UUT is the scheduler's policy. As it can be observed for RR under good channel quality conditions, the UUT is hardly assigned up to 40 codes every 10 TTIs, i.e. 4 codes per TTI in average. On the other hand, the value of this parameter for MaxC/I under similar channel quality conditions usually reaches the average of 5 codes per TTI. This is due to the fact that users experiencing favorable channel conditions tend to monopolize resources when the MaxC/I algorithm is applied.

To analyze the impact of the instantaneously experienced throughput on the user perceived QoE, Figure 18 and Figure 19 show some screenshots of the transmitted video sequence during three different phases. The first phase corresponds to the first handover between cells (poor channel quality conditions) at time instants around 65 sec, whose consequences are observed some seconds later on. The second phase corresponds to time instants around 150 sec in which the experienced channel quality is favorable. Finally, the third phase corresponds to the second handover between cells at time instants around 195 sec. During the first phase, some degradation is observed in the image quality. This is due to the throughput reduction experienced as a result of the poor channel quality conditions. Although this behavior is observed for both scheduling algorithms, a more serious degradation is observed for MaxC/I since the number of transmission requests rejected during the first phase is higher in the MaxC/I case than in the RR case (top-right side of the figures). During the second phase, the video sequence is transmitted in both cases without noting any distortion since the channel quality conditions are favorable. The higher peak data-rates offered by the MaxC/I scheduler during this phase are not reflected in the instantaneous image quality. In fact, the image quality during the second phase for the RR scheduler is as good as for the MaxC/I scheduler. The difference becomes apparent, however, during the third phase, when the second handover occurs. As a result of the higher data-rate experienced during the second phase when the MaxC/I criterion is applied, the UUT's reception buffer is almost full by the time the handover occurs. Thus, although some transmission requests are rejected during the third phase, the sequence can be reproduced without loss of continuity and no degradation is observed. In the RR case, the data-rates experienced during the second phase are not as high as in the MaxC/I case, and the UUT's buffer is not so full by the time the second handover occurs. As a result, there exist some time instants in which the UUT's buffer gets empty, which indeed leads to the image degradation observed in the third screenshot of Figure 19.

This case study has shown that, although the testbed implements a complete B3G IP-based heterogeneous wireless network, its detailed implementation makes it possible to perform specific studies related to low level features for particular components and technologies, without loss of detail with respect to specific-purpose evaluation tools.

8.2. Case Study II: Mobility Management and QoE

This case study is devoted to illustrate the mobility management strategies implemented in the testbed and to show the tesetbed's capabilities in assessing the user QoE. In this example, the UUT requests a streaming session with a guaranteed bit-rate of 192 kbps and makes use of real client applications to watch the streamed movie. Concretely, Darwin streaming server is run on the applications server, which contains media of different bit-rates and codecs including video and audio. For all the trials presented, a 128 kbps video sequence of approximately 120 seconds encoded with a H.264 variable bit-rate video codec is used. This video sequence is requested by a VLC player running in the applications client machine.

In all this example, the UUT moves within an 8 km × 4 km service area with 13 UTRAN and 13 GERAN co-located base stations and 6 WLAN hot-spots. However, GERAN is not considered here because of the limited capabilities offered by this RAN for a proper QoS provision under streaming services. The desired HOs in this study are produced by properly defining the UUT's trajectory between base stations and CRRM technology preference weights for RAN selection algorithms [32]. To evaluate intra-IR VHO procedures, the considered scenario assumes that UTRAN and WLAN are both attached to IR1. For the rest of HO procedures, UTRAN is attached to IR1 whereas WLAN is attached to IR2. Apart from the UUT, a total of 1000 emulated users are uniformly distributed over the service area: 500 conversational, 300 interactive and 200 streaming users. The UUT moves along the scenario and requests streaming sessions with guaranteed bit-rates of 192 kbps in downlink.

Different conditions have been evaluated depending on the type of HO evaluated in each case. To test HHOs, a periodic HHO between two UTRAN base stations is considered.

For VHO experiments, periodic VHOs between WLAN and UTRAN are forced. In the case of intra-IR VHO, the transfer policy effectiveness is compared to the case where no advanced policy is used. VHOs include IR change in the case of inter-IR VHO. Notice that in such a case the CN MPLS tunnel switching is triggered once the MN entity detects an IR change based on the received RAs. Three different RA periods have been tested, namely 1, 5 and 10 seconds.

Figure 20 depicts the average packet loss measured at the UUT PC for the different HO types studied in this example. For testing these values, a 128 kbps constant bit-rate UDP downlink stream is sent from the applications server to the applications client. Each value was obtained by averaging the statistics collected during 30 minute periods, during which around 100 HOs occurred. As it can be appreciated, no packet losses are observed for HHOs. In the case of intra-IR VHO, the experienced packet losses are practically negligible when the transfer policy is enabled, but appreciable when the transfer policy is disabled due to some packets being addressed to the old RAN but not transferred to the new RAN. Finally, a comparison of inter-IR VHOs with and without HO preparation is shown. In both cases, it is observed that the average packet loss is greater than in the intra-IR VHO case because of the data path switching mechanisms in the RAN and CN parts. When HO preparation is disabled, packet losses increase with the RA period due to longer periods of misalignment between the paths through the RAN and CN domains. As a result, the greatest packet loss is measured for a 10s RA period with no HO preparation. On the other hand, when the HO preparation mechanism is enabled, the inter-IR tunnel allows maintaining packet losses below 2.5% regardless of the RA period.

In order to qualitatively validate the considered HO procedures, the testbed offers the possibility to visualize in real-time the statistics of the different modules in execution. Figure 21 shows an example of the testbed statistics when intra-IR VHOs occur during one of the packet loss experiments explained above. Graphs (a) and (b) depict the current RAN the UUT is attached to (UTRAN = 0 and WLAN = 2). Therefore, the instants when a VHO occurs can be dynamically appreciated in the real-time statistics. Graphs (c) and (d) show the current IR in use. It can be observed that in this trial there is no IR change every time a VHO is performed. Finally, graphs (e) and (f) represent the amount of bytes transmitted to the UUT. The left-hand side graphs present the case where the transfer policy was disabled. In this case, it can be observed that significant throughput cuts are experienced in every VHO, which may lead to important packet losses and unacceptable delays. On the other hand, no throughput cut-offs are perceived in the right-hand side, where the transfer policy was enabled. This kind of real-time statistics can be very helpful for fast and qualitative validations.

The user QoE is shown in Figure 22, expressed in terms of Mean Opinion Score (MOS). These values have been computed by averaging 10 repetitive tests for each HO type and using a full-reference model-based objective metric [64] for the QoS evaluation based on ITU recommendations [40]. This kind of evaluation methods compares a reference non-degraded sample of the video sequence to a degraded sample obtained at the output of the system (e.g., after passing through the testbed) in order to provide a quantitative satisfaction level. This metric aims at expressing the subjective score that human beings would give to the experiment. Satisfaction levels are numeric values in the scale from 1 to 5, where the satisfaction level 5 corresponds to a perfect video quality (e.g., the transmitted video sequence is perceived by the user as not degraded at all), while a score of 1 means complete loss of information. It is worth noting that these methods rarely provide a satisfaction level of 5 since human perception is reluctant to assign the maximum score (i.e., perfect perceived quality) even if the compared videos are identical. As it can be observed, in general greater MOS values are obtained for lower packet losses. Again, the HO preparation mechanism considerably improves the QoE metric obtained and, independently of the RA period, the streaming session quality is considered good by the UUT. Thus, the HO preparation mechanism adds robustness to the user session and helps preserving the user experience.

Figure 23 illustrates the system behavior during an inter-IR VHO with and without HO preparation. These results have been obtained for a RA period equal to 10 seconds. The left-hand side of Figure 23 shows the statistics without HO preparation, whereas the right-hand side plots the statistics with this functionality enabled. Graphs (a) and (b) depict the instants when the MN triggers the MPLS tunnel switching, while graphs (c) and (d) represent the instants where VHOs are performed. As it can be observed, the MN realizes that a MPLS tunnel switching needs to be triggered some time after the VHO is performed. During a given interval, the RAN and CN paths are misaligned since the RAN part is attached to the new IR while the

CN part is still delivering packets to the old IR. As a result, some disruptions in the transmitted bytes are observed in Figure 23 when the HO preparation mechanism is disabled. On the other hand, when enabling this mechanism no throughput disruptions are perceived during the VHO execution as a result of the inter-IR tunnel previously established. The corresponding user perceived QoE for both cases is shown in Figure 24, where several snapshots of the video sequence received at the applications client are shown. As it can be appreciated, without HO preparation the video sequence freezes during the VHO execution due to the significant level of packet losses caused by the misalignment between the RAN and CN parts. With HO preparation enabled, the video sequence is played normally since no downstream packets are lost during the VHO execution.

This case study has shown how the testbed can be employed to assess different mobility management strategies during mobility events such as horizontal and vertical handovers. Moreover, the impact of each one of the considered mobility management strategies on the user QoE has been analyzed as well.

8.3. Case Study III: e2e QoS and System Reconfiguration

This case study presents an example with complete e2e system reconfigurations, involving all the testbed entities in order to provide adequate QoS levels. A real user moving under constant UTRAN coverage and occasional WLAN coverage is considered, as shown in Figure 25. The UUT performs two requests, first a FTP file download, followed by a video streaming sequence. During the session lifetime, a set of actions involving QoS renegotiation procedures are performed. The points where such actions are executed are shown in Figure 25, and the corresponding time instances are detailed in Table 2. Figure 26 shows the UUT throughput while connected to UTRAN, WLAN, and the overall throughput at the user terminal, as well as the RAN the UUT is connected to at any time (UTRAN = 0 and WLAN = 2) and the overall throughput in the IRs. The user QoE is depicted in Figure 27.

In this experiment, QoS renegotiation procedures are triggered five times: two times by the QoS Client for session modification (actions 3 and 4), two times by the CRRM (actions 5 and 8), and once by the BB (action 10).

The first negotiation triggered by the user (action 3, session throughput modification) includes the communication between QoS Client and WQB/CRRM. The BB is not involved in this case since the user remains connected to the same RAN and thus to the same IR. Therefore, the CRRM modifies the session throughput locally inside the same RAN (UTRAN) with no IR change. However, when the user decides to request a service class modification (action 4), the considered network polices result in a VHO from UTRAN to WLAN, requested by the CRRM module. Since WLAN is attached to a different IR, such VHO implies an IR change and in this case the BB is also involved.

The first reconfiguration process initiated by the CRRM is due to the loss of WLAN coverage, which forces a VHO to UTRAN (action 5). The second one is triggered when the WLAN coverage is detected again; in this case, the VHO from UTRAN to WLAN is triggered due to the CRRM preferences (action 8). In both cases, the VHO implies an IR change. The BB is therefore contacted to check if the reconfiguration is possible. Notice that in the first case the session would be dropped in case of reconfiguration rejection from the BB side, while in the second case they session would stay active on the same RAN (i.e., the VHO would not be performed).

The reconfiguration triggered by the BB (action 10) is a consequence of the congestion situation caused by the temporary load increase experienced at IR2 (where the UUT's current RAN, i.e. WLAN, is attached). Such load increase was artificially induced by injecting additional traffic to IR2. This example however serves as a proof of concept, demonstrating the possibility that the BB detects a congestion situation in the CN and suggests a VHO in the RAN part (from WLAN/IR2 to UTRAN/IR1) with the aim of preserving the QoS level. Therefore, this is an example of a peculiar case where a VHO in the RAN part is triggered from the CN part as a result of a congestion situation in order to preserve the e2e QoS level. This example illustrates the complexity of the e2e QoS and mobility mechanisms implemented in the testbed, their interaction, and the testbed's ability in handling and emulating such kind of complex situations within the local domains and from an e2e perspective.

9. CONCLUSIONS

This work has presented the real-time emulation platform for all-IP beyond 3G heterogeneous wireless networks that was developed in the framework of the European IST AROMA research project (http://www.aroma-ist.upc.edu). The main objective of the AROMA testbed is to provide a highly accurate and realistic framework where the performance of algorithms, policies, protocols, services and applications for a complete heterogeneous wireless network can be fully assessed and evaluated before bringing them to a real system. Besides the important but often isolated simulators and specific targeted research testbeds, it is also necessary to provide the research community with an environment that reflects a complete real world heterogeneous wireless network. In this context, this work has provided a detailed description of an ambitious and sophisticated testbed that enables the real-time emulation at the packet/slot level of an entire all-IP heterogeneous wireless network with a high level of implementation detail. This work provides an in-depth description of the presented tool, emphasizing many interesting implementation details and lessons learned during its development that may result helpful to other researchers and system engineers in the development of similar tools. Several case studies illustrating the potentials and capabilities of the presented platform have been presented as well. As a part of the future work, the potential addition of novel radio technologies such as LTE and cognitive radio in the radio access network as well as the extension of the number of nodes in the core network are envisaged.

ACKNOWLEDGMENTS

This work was performed in the framework of the IST AROMA project (http://www.aroma-ist.upc.edu), which was partly funded by the European Community (Ref. IST-4-027567).

REFERENCES

[1] M. Mouly and M.-B. Pautet, "The GSM System for Mobile Communications," Telecom Publishing (1992).

[2] T. Halonen, J. Romero and J. Melero, "GSM, GPRS and EDGE: Evolution towards 3G/UMTS," 2nd edition, John Wiley & Sons (2003).

[3] H. Holma and A. Toskala, "WCDMA for UMTS: Radio Access for Third Generation Mobile Communications," John Wiley & Sons (2004).

[4] H. Holma and A. Toskala, "HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications," John Wiley & Sons (2006).

[5] H. Holma and A. Toskala, "WCDMA for UMTS: HSPA Evolution and LTE," John Wiley & Sons (2007).

[6] H. Holma and A. Toskala, "LTE for UMTS: OFDMA and SC-FDMA Based Radio Access," John Wiley & Sons (2009).

[7] J. G. Andrews, A. Ghosh and R. Muhamed, "Fundamentals of WiMAX: Understanding Broadband Wireless Networking," Prentice Hall (2007).

[8] I. Oppermann, M. Hämäläinen and J. Iinatti, "UWB: Theory and Applications," John Wiley & Sons (2004)

[9] U. Reimers, "DVB: The Family of International Standards for Digital Video Broadcasting," Springer (2004).

[10] M. S. Gast, "802.11 Wireless Networks," O'Reilly (2002).

[11] J. Bray and C. F. Sturman, "Bluetooth: Connect without Cables," Prentice Hall (2002).

[12] A. Tölli, P. Hakalin and H. Holma, "Performance Evaluation of Common Radio Resource Management (CRRM)," in proceedings of the IEEE International Conference on Communications (ICC 2002), April 2002, vol. 5, pp. 3429 – 3433.

[13] IST AROMA Project (Advanced Resource Management Solutions for Future All IP Heterogeneous Mobile Radio Environments), 6th Framework Program of the European Community, http://www.aroma-ist.upc.edu.

[14] F. Bernardo, N. Vučević, Ł. Budzisz and A. Umbert, "A beyond 3G real-time testbed for an all-IP heterogeneous network," in proceedings of the 5th ACM International Workshop on Mobility Management and Wireless Access (MobiWAC 2007), October 2007, pp. 50-59.

[15] A. Umbert, M. López-Benítez, F. Bernardo, N. Vučević, R. Azevedo and A. Oliveira, "The real-time AROMA testbed for all-IP heterogeneous wireless access networks," in proceedings of the 4th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities (TRIDENTCOM 2008), March 2008, 10 pp.

[16] A. Umbert, M. López-Benítez, N. Vučević and F. Bernardo, "A real-time platform for end-to-edge QoS evaluation in heterogeneous networks," in proceedings of the ICT Mobile and Wireless Communications Summit (ICT MobileSummit 2008), June 2008, 8 pp.

[17] The Bay Area Research Wireless Access Network (BARWAN) project. Available at: http://bnrg.eecs.berkeley.edu/~randy/Daedalus/BARWAN/BARWAN_index.html

[18] B. Ionescu, M. Ionescu, S. Veres, D. Ionescu, F. Cuervo and M. Luiken-Miller, "A Testbed and Research Network for Next Generation Services over Next Generation Networks," in proceedings of the First International Conference on Testbeds and Research Infrastructures for the DEvelopment of NeTworks and COMmunities (TRIDENTCOM 2005), February 2005, 10 pp.

[19] I. Carreras, R. Grasso, C. Kiraly, S. Pera, H. Woesner, Y. Ye and C. A. Szabó, "Design Considerations on the CREATE-NET Testbed," in proceedings of the First International Conference on Testbeds and Research Infrastructures for the DEvelopment of NeTworks and COMmunities (TRIDENTCOM 2005), February 2005, 10 pp.

[20] S. Guruprasad, R. Ricci and J. Lepreau, "Integrated Network Experimentation using Simulation and Emulation," in proceedings of the First International Conference on Testbeds and Research Infrastructures for the DEvelopment of NeTworks and COMmunities (TRIDENTCOM 2005), February 2005, 9 pp.

[21] S. Hu, G. Wu, Y. Liang Guan, C. Look Law, Y. Yan and S. Li, "Development and Performance Evaluation of Mobile WiMAX Testbed," Mobile WiMAX Symposium, March 2007, 4 pp.

[22] N. Scalabrino, F. De Pellegrini, R. Riggio, A. Maestrini, C. Costa and I. Chlamtac, "Measuring the Quality of VoIP Traffic on a WiMAX Testbed," in proceeding of the 3rd International Conference on Testbeds and Research Infrastructure for the Development of Networks and Communities (TRIDENTCOM 2007), 21-23 May 2007, 10 pp.

[23] M. Takai, R. Bagrodia, M. Gerla, B. Daneshrad, M. Fitz, M. Srivastava, E. Belding-Royer, S. Krishnamurthy, M. Molle, P. Mohapatra, R. Rao, U. Mitra, C.-C. Shen and J. Evans, "Scalable Testbed for Next-Generation Wireless Network Technologies," in proceedings of the First International Conference on Testbeds and Research Infrastructures for the DEvelopment of NeTworks and COMmunities (TRIDENTCOM 2005), February 2005, 10 pp.

[24] N. Brownlee, P. Christ, J. Jaehner, Y. Liang, K. Srinivasm and J. Zhou, MobyDick FlowVis Using NeTraMet for distributed protocol analysis in a 4G network environment," in proceedings of the 3rd IEEE Workshop on IP Operations and Management, 2003. (IPOM 2003), Oct. 2003, 6 pp.

[25] DAIDALOS - Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimised personal Services, http://www.ist-daidalos.org/

[26] F. Steuer, M. Elkotob, S. Albayrak and A. Steinbach, "Testbed for mobile network operator scenarios," in proceedings of the 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TRIDENTCOM 2006), March 2006, 10 pp.

[27] A. Stone, "Investigating wireless networks with WHYNET," IEEE Computer Society, Vol. 5, No. 4, April 2004, 6 pp.

[28] M. López-Benítez, Ł. Budzisz, A. Umbert, N. Vučević and F. Bernardo, "A real-time testbed for heterogeneous wireless networks," demonstration performed at the 1st IEEE

International Symposium on Wireless Vehicular Communications (WiVeC 2007), September – October 2007, 2 pp.

[29] F. Bernardo, A. Umbert, M. López-Benítez and N. Vučević, “Advanced and versatile real-time emulation platform for heterogeneous radio access systems,” demonstration performed at the Eleventh ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM 2008), appears in proceedings of the Third ACM International Workshop on Performance Monitoring, Measurement, and Evaluation of Heterogeneous Wireless and Wired Networks (PM2HW2N 2008), October 2008, pp 180-185.

[30] M. López-Benítez, F. Bernardo, N. Vučević and A. Umbert, “Real-time HSPA emulator for end-to-edge QoS evaluation in all-IP beyond 3G heterogeneous wireless networks,” in proceedings of the 1st International Workshop on the Evaluation of Quality of Service through Simulation in the Future Internet (QoSim 2008), March 2008, 12 pp.

[31] A. Umbert, Ł. Budzisz, N. Vučević and F. Bernardo, “An all-IP heterogeneous wireless testbed for RAT selection and e2e QoS evaluation,” in proceedings of the 1st International Workshop on Broadband Wireless Access (BWA 2007), September 2007, pp. 310-315.

[32] M. López-Benítez, N. Vučević, F. Bernardo and A. Umbert, “Real-time evaluation of radio access technology selection policies in heterogeneous wireless systems: The AROMA testbed approach,” in proceedings of the Eleventh ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM 2008), October 2008, pp 294-302.

[33] N. Vučević, F. Bernardo, A. Umbert and Ł. Budzisz, “Evaluation of perceived QoS with multimedia applications in a heterogeneous wireless network,” in proceedings of the 4th IEEE International Symposium on Wireless Communication Systems (ISWCS 2007), October 2007, pp. 102-106.

[34] F. Bernardo, N. Vučević, A. Umbert and M. López-Benítez, “Quality of experience evaluation under QoS-aware mobility mechanisms,” in proceedings of the 14th European Wireless Conference 2008 (EW 2008), June 2008, 7 pp.

[35] N. Vučević, F. Bernardo, A. Umbert and M. López-Benítez, "End-to-edge QoS across heterogeneous wireless and wired domains," in proceedings of the Fifth International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine 2008), July 2008, 7 pp.

[36] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, "An Architecture for Differentiated Services," RFC 2475, December 1998.

[37] E. Rosen, A. Viswanathan and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, January 2001.

[38] ITU-R Recommendation BS.1387, "Method for Objective Measurements of Perceived Audio Quality," November 2001.

[39] ITU-R Recommendation BT.1683, "Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference," June 2004.

[40] ITU-T Recommendation J.144, "Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference," March 2004.

[41] ITU-T Recommendation P.862, "Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs," February 2001.

[42] IEEE Std 1003.3-1991, "IEEE Standard for Information Technology-Test Methods for Measuring Conformance to POSIX".

[43] A. Gelonch and R. Ferrús (editors), "Implemented Testbed: Subsystems and Modules," EVERST Project, Deliverable D12, January 2005, Annex B, pp. 97-146, available at http://www.everest-ist.upc.es.

[44] 3GPP TR 101 112, "Universal Mobile Telecommunications System (UMTS); Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03 version 3.2.0)," v3.2.0, April 1998.

[45] M. Gudmundson, "Correlation Model for Shadow Fading in Mobile Radio Systems," Electronics Letters, vol. 27, nº 23, pp. 2145-2146, November 1991.

[46] 3GPP TS 43.129, “Packet-switched handover for GERAN A/Gb mode; Stage 2 (Release 7),” v7.2.0, May 2007.

[47] 3GPP TR 25.922, “Radio resource management strategies (Release 7),” v7.1.0, March 2007.

[48] 3GPP TS 22.129, “Handover requirements between UTRAN and GERAN or other radio systems (Release 8),” v8.1.0, December 2007.

[49] P.Brady, “A Model for ON-OFF Speech Patterns in Two-Way Communications,” Bell System Technology Journal, vol. 48, pp. 2245-2472, September 1969.

[50] J. Laiho, A. Wacker and T. Novosad, “Radio Network Planning and Optimisation for UMTS,” Wiley (2002).

[51] H. Holma and A. Toskala, “WCDMA for UMTS: Radio Access for Third Generation Mobile Communications,” Wiley (2004).

[52] A. Umbert (editor), “Testbed Specification,” AROMA Project, Deliverable D07, July 2006, available at http://www.aroma-ist.upc.edu.

[53] B. Davie, A. Charny, J.C.R. Bennett, K. Benson, J.Y. Le Boudec, W. Courtney, S. Davari, V. Firoiu and D. Stiliadis, “An Expedited Forwarding PHB (Per-Hop Behavior),” RFC 3246, March 2002.

[54] J. Heinanen, F. Baker, W. Weiss and J. Wroclawski, “Assured Forwarding PHB Group,” RFC 2597, June 1999.

[55] T. M. T. Nguyen, N. Boukhatem, Y. G. Doudane and G. Pujolle, “COPS-SLS: A Service Level Negotiation Protocol for the Internet,” IEEE Communications Magazine, May 2002, vol. 40, nº 5, pp.158-165.

[56] D. P. Audsin (editor), “Integrated Testbed,” AROMA Project, Deliverable D16, July 2007, available at http://www.aroma-ist.upc.edu.

[57] C. Perkins, “IP Mobility Support for IPv4,” RFC 3344, August 2002.

[58] A. T. Campbell, J. Gómez, S. Kim, C.-Y. Wan, Z. R. Turanyi and A. G. Valkó, “Comparison of IP Micromobility Protocols,” IEEE Wireless Communications, February 2002, vol. 9, nº 1, pp. 72-82.

[59] H. Soliman, C. Castelluccia, K. El Malki and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," RFC 4140, August 2005.

[60] C. Boukis, N. Georganopoulos and H. Aghvami, "A Hardware Implementation of BCMP Mobility Protocol for IPv6 Networks," in proceedings of the IEEE Global Telecommunications Conference (GLOBECOM 2003), December 2003, vol. 6, pp. 3083-3087.

[61] R. Ramjee, K. Varadhan, L. Salgarelli, S. R. Thuel, S.-Y. Wang and T. La Porta, "HAWAII: A Domain-based Approach for Supporting Mobility in Wide-Area Wireless Networks," IEEE/ACM Transactions on Networking, June 2002, vol. 10, nº 3, pp. 396-410.

[62] A. G. Valkó, "Cellular IP: A New Approach to Internet Host Mobility," ACM SIGCOMM Computer Communication Review, January 1999, vol. 29, nº 1, pp. 50-65.

[63] A. Umbert (editor), "Trial results and algorithm validation," AROMA Project, Deliverable D20, October 2008, available at http://www.aroma-ist.upc.edu.

[64] G.-M. Muntean, P. Perry, L. Murphy, "Objective and Subjective Evaluation of QOAS Video Streaming over Broadband Networks," IEEE Transactions on Network and Service Management, November 2005, vol. 2, nº 1, pp. 19-28.

TABLES

Table 1. Some applications employed in the testbed [1].

End-to-end service	End-to-end application		Capturing software
	Server	**Client**	
Web browsing	Apache HTTP Server	Internet Explorer	—
		Mozilla Firefox	
Video streaming	Darwing Streaming Server	QuickTime Pro	Camtasia Studio Recorder
		VLC	
Video conference	VIC		
	NetMeeting		
Audio conference	RAT		Microsoft Sound Recorder
	NetMeeting		

Table 2. Actions performed in the case study.

	Action	Comment	Time
1	User starts session	Interactive class, 64 kbps in downlink	
2	User starts file download	FTP client/server	~190s
3	User modifies session throughput	Increase downlink bandwidth to 96 kbps	~220s
4	User modifies service class to streaming (same throughput)	Results in VHO due to the CRRM preferences for the streaming class	~240s
5	CRRM initiates VHO	Due to the loss of coverage in the current RAN	~270s
6	User stops file download	FTP client/server	~310s
7	User starts video streaming	Video 64 kbps and audio 24 kbps	~380s
8	CRRM initiates VHO	Due to CRRM preferences	~440s
9	Congestion in IR2 (current AR)	Additionally generated traffic	~460s
10	BB suggests reconfiguration	Results in VHO due to the congestion detection	~540s
11	User stops session	Resources are released	

[1] Apache HTTP Server, web site: http://httpd.apache.org/
Internt Explorer, web site http://www.microsoft.com/windows/products/winfamily/ie/default.mspx
Mozilla Firefox, web site: http://www.mozilla.com/firefox
Darwin Streaming Server, web site: http://developer.apple.com/opensource/server/streaming/index.html
QuickTime Pro, web site: http://www.apple.com/quicktime
VLC media player, web site: http://www.videolan.org/vlc
Video Conferencing Tool (VIC), web site: http://www-nrg.ee.lbl.gov/vic
Robust Audio Tool (RAT), web site: http://www-mice.cs.ucl.ac.uk/multimedia/software/rat
NetMeeting, web site: http://www.microsoft.com/windows/netmeeting
Camtasia Studio, web site: http://www.techsmith.com/camtasia.asp

FIGURES

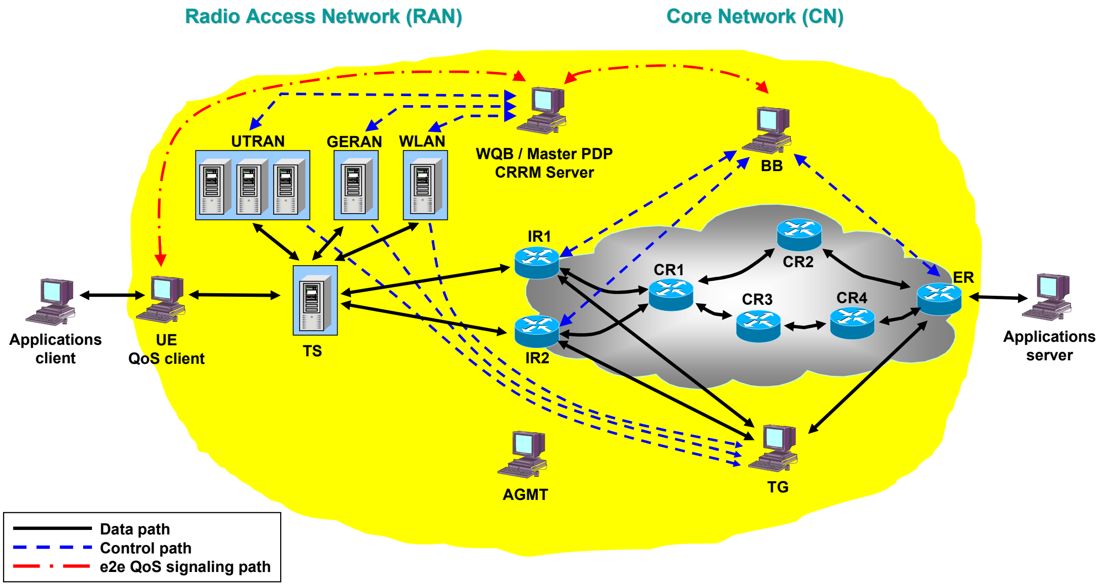


Figure 1. The AROMA testbed architecture.

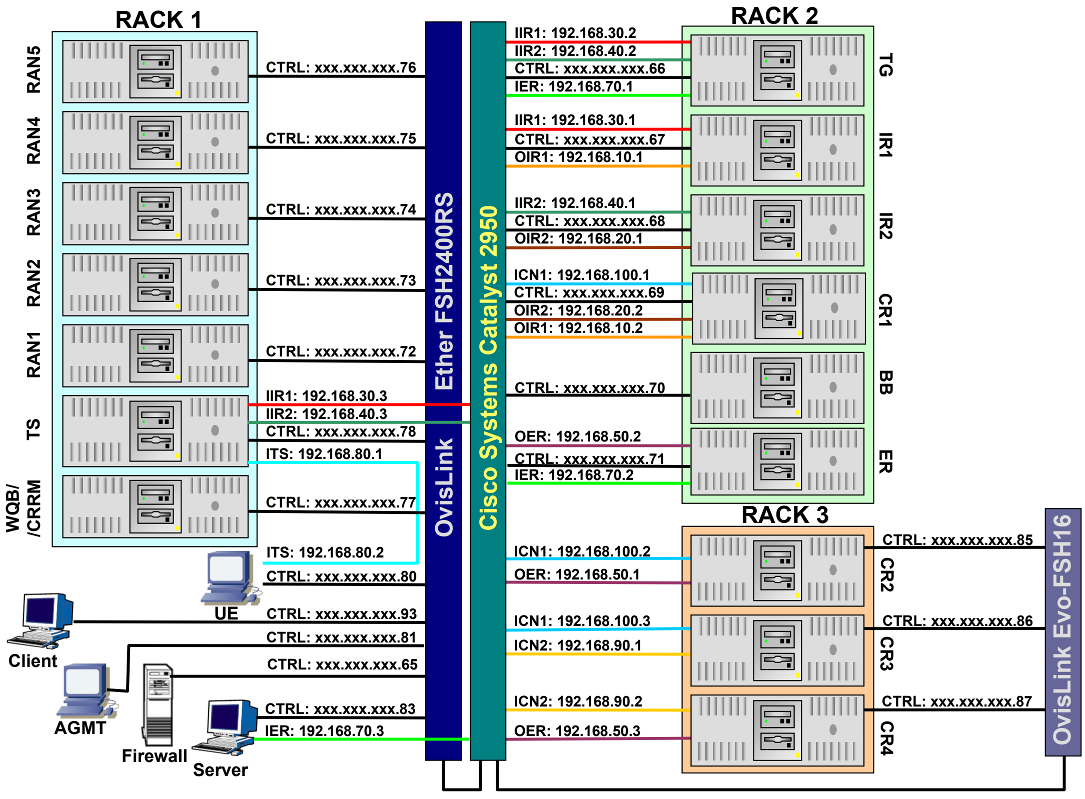


Figure 2. The AROMA testbed hardware platform.

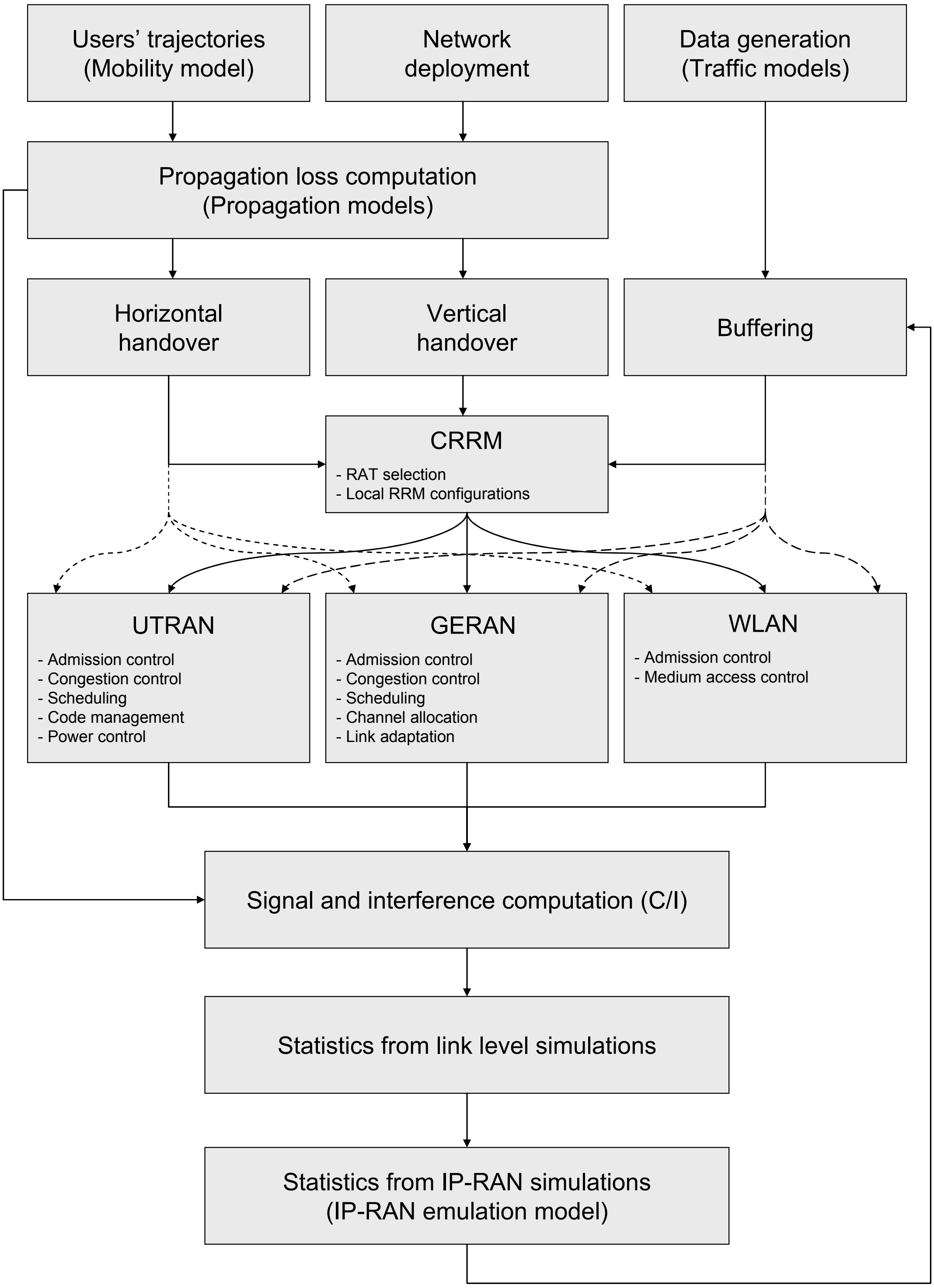


Figure 3. Functional RANE architecture.

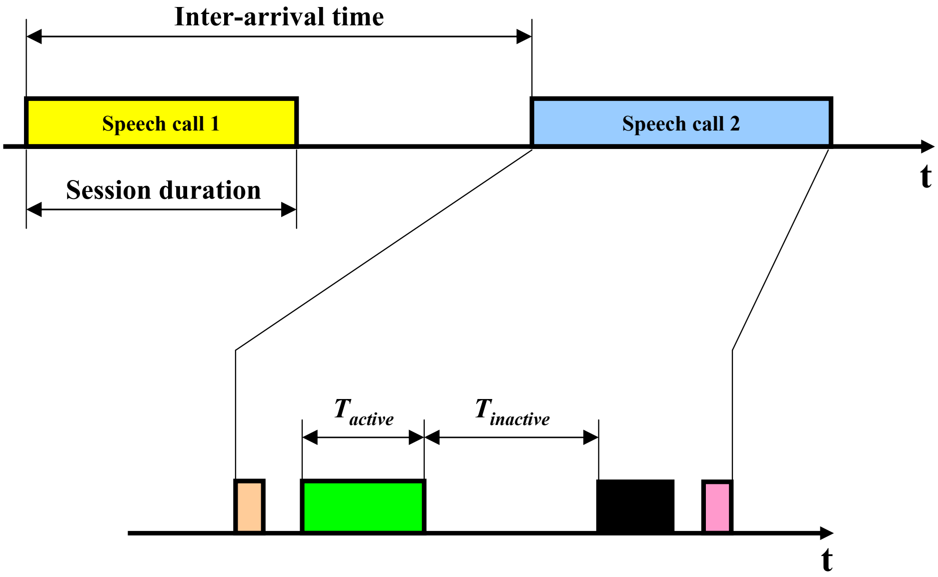


Figure 4. Voice model.

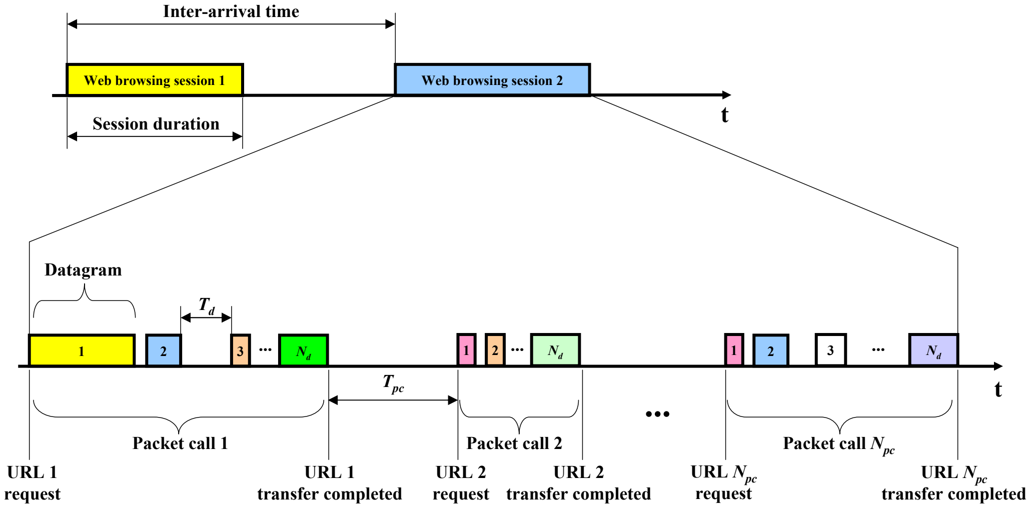


Figure 5. Web browning model.

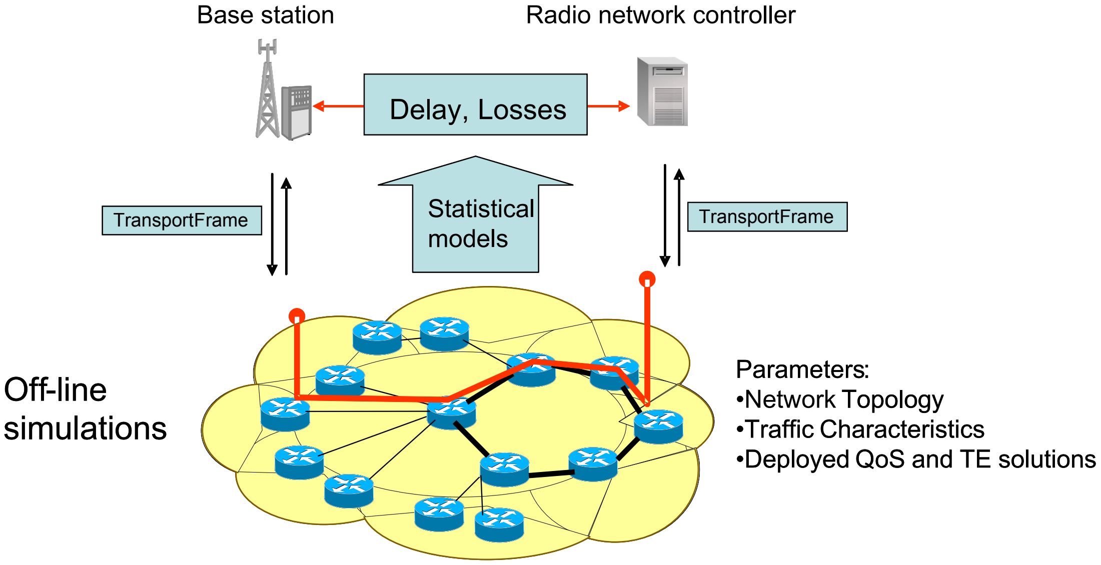


Figure 6. IP-RAN model.

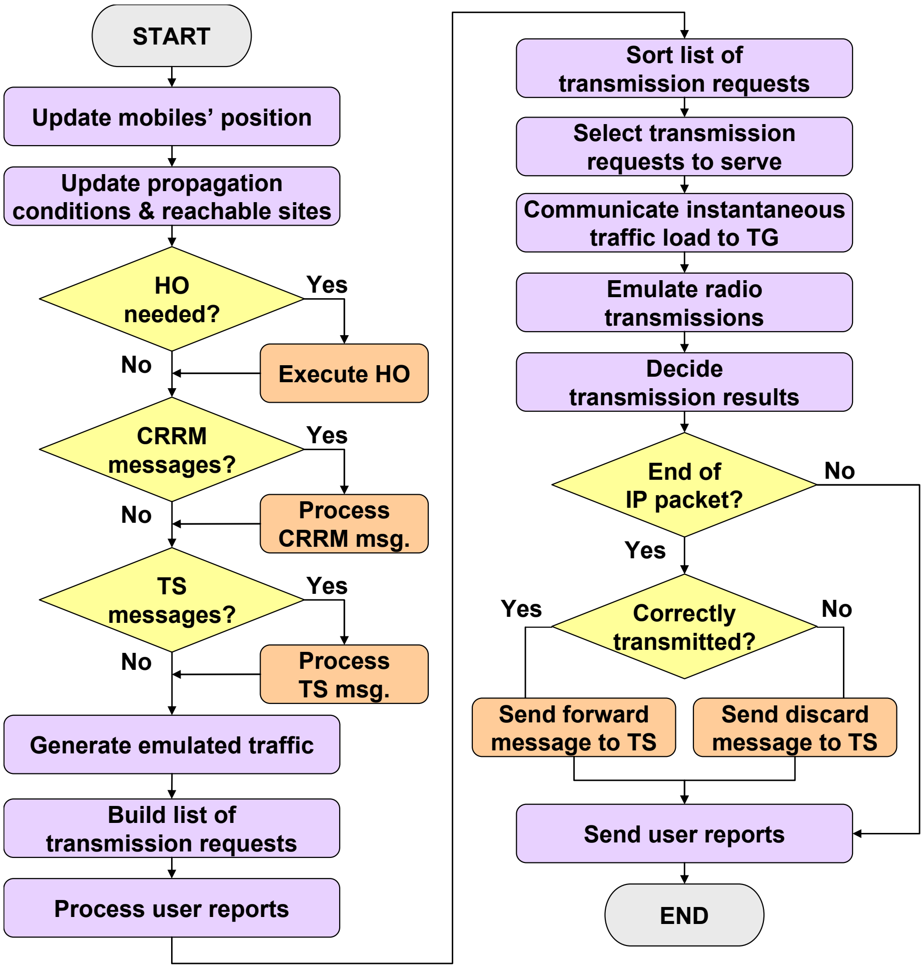

START
Update mobiles' position
Update propagation conditions & reachable sites
HO needed?
Yes
Execute HO
No
CRRM messages?
Yes
Process CRRM msg.
No
TS messages?
Yes
Process TS msg.
No
Generate emulated traffic
Build list of transmission requests
Process user reports
Sort list of transmission requests
Select transmission requests to serve
Communicate instantaneous traffic load to TG
Emulate radio transmissions
Decide transmission results
End of IP packet?
No
Yes
Correctly transmitted?
Yes
No
Send forward message to TS
Send discard message to TS
Send user reports
END

Figure 7. RANEs execution loop.

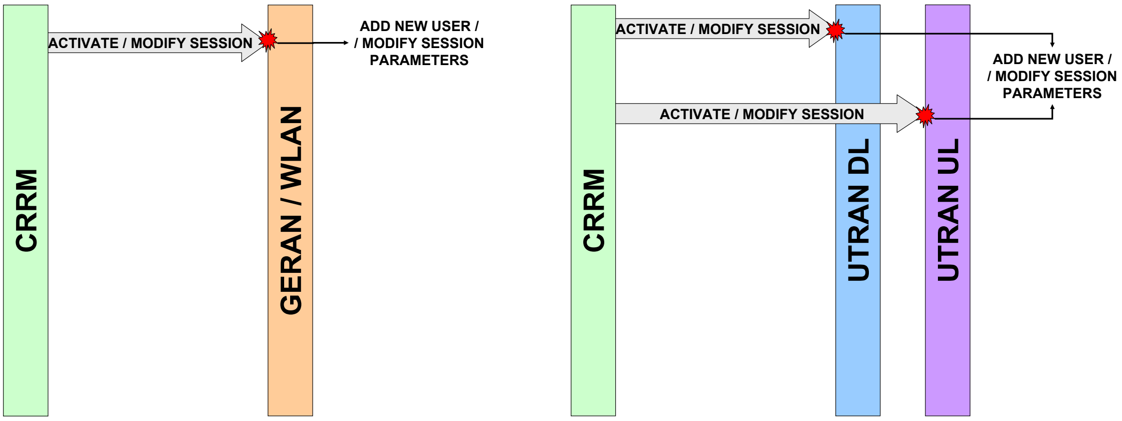

CRRM
ACTIVATE / MODIFY SESSION
GERAN / WLAN
ADD NEW USER / / MODIFY SESSION PARAMETERS
CRRM
ACTIVATE / MODIFY SESSION
ACTIVATE / MODIFY SESSION
UTRAN DL
UTRAN UL
ADD NEW USER / / MODIFY SESSION PARAMETERS

Figure 8. Session activation/modification.

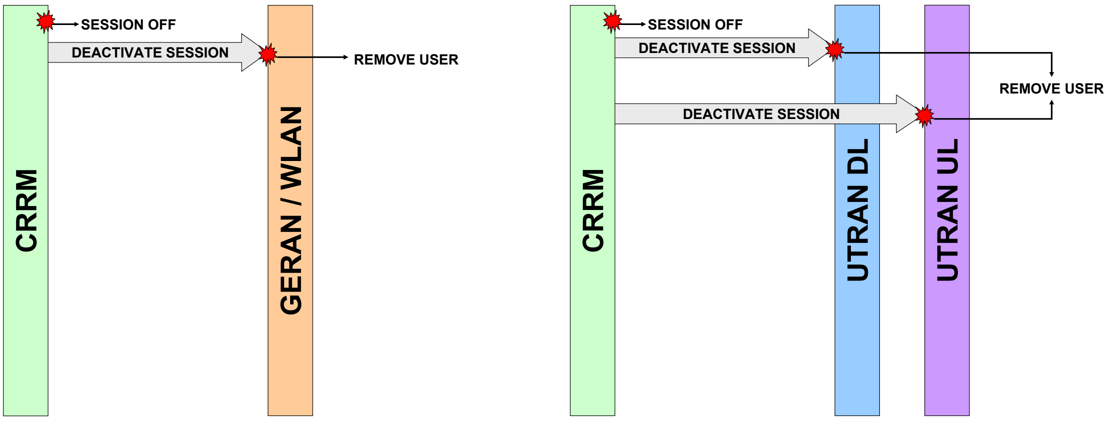


Figure 9. Session deactivation from CRRM.

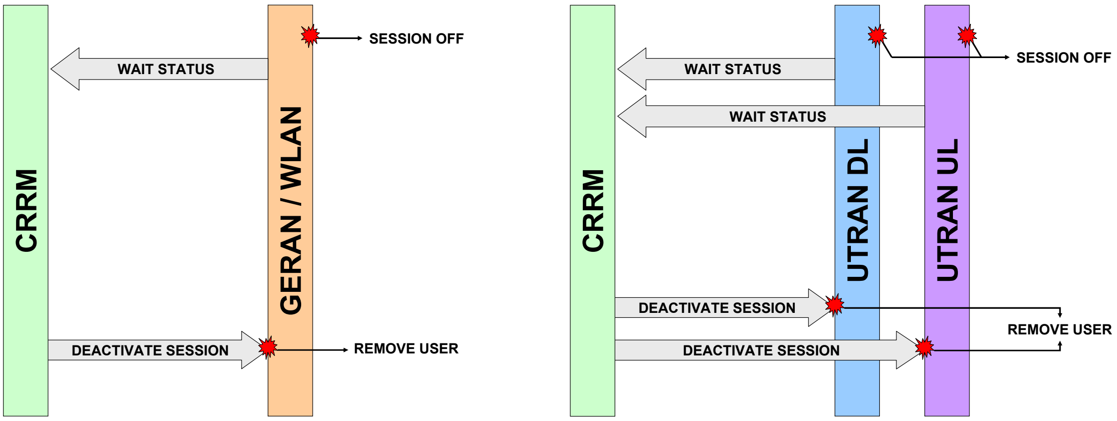


Figure 10. Session deactivation from RANEs.

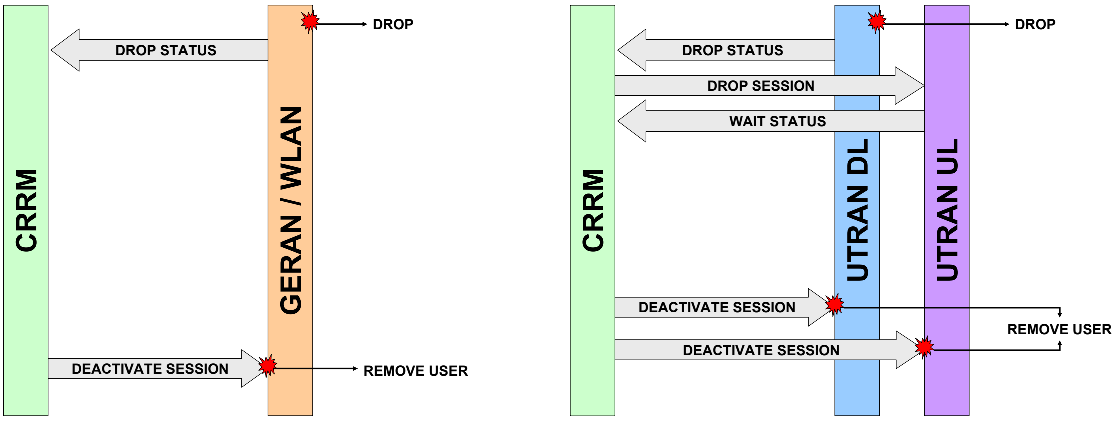


Figure 11. Session dropping.

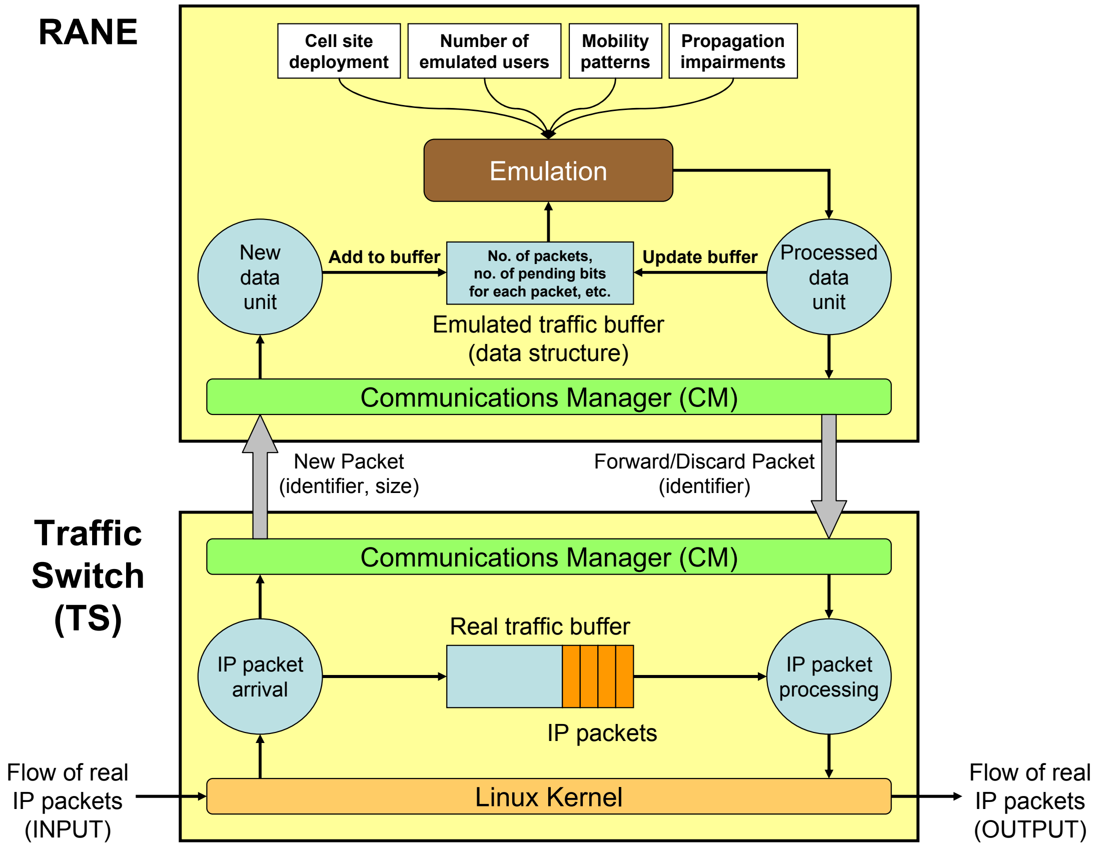

RANE
Cell site deployment
Number of emulated users
Mobility patterns
Propagation impairments
Emulation
New data unit
Add to buffer
No. of packets, no. of pending bits for each packet, etc.
Update buffer
Processed data unit
Emulated traffic buffer (data structure)
Communications Manager (CM)
New Packet (identifier, size)
Forward/Discard Packet (identifier)
Traffic Switch (TS)
Communications Manager (CM)
Real traffic buffer
IP packet arrival
IP packet processing
IP packets
Flow of real IP packets (INPUT)
Linux Kernel
Flow of real IP packets (OUTPUT)

Figure 12. Interaction RANEs-TS.

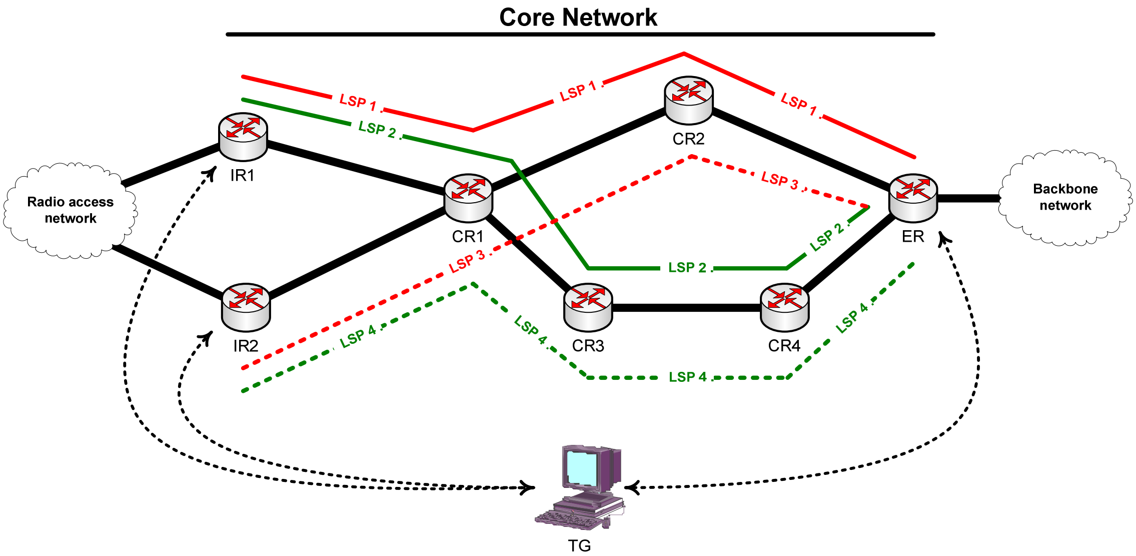


Figure 13. CN topology.

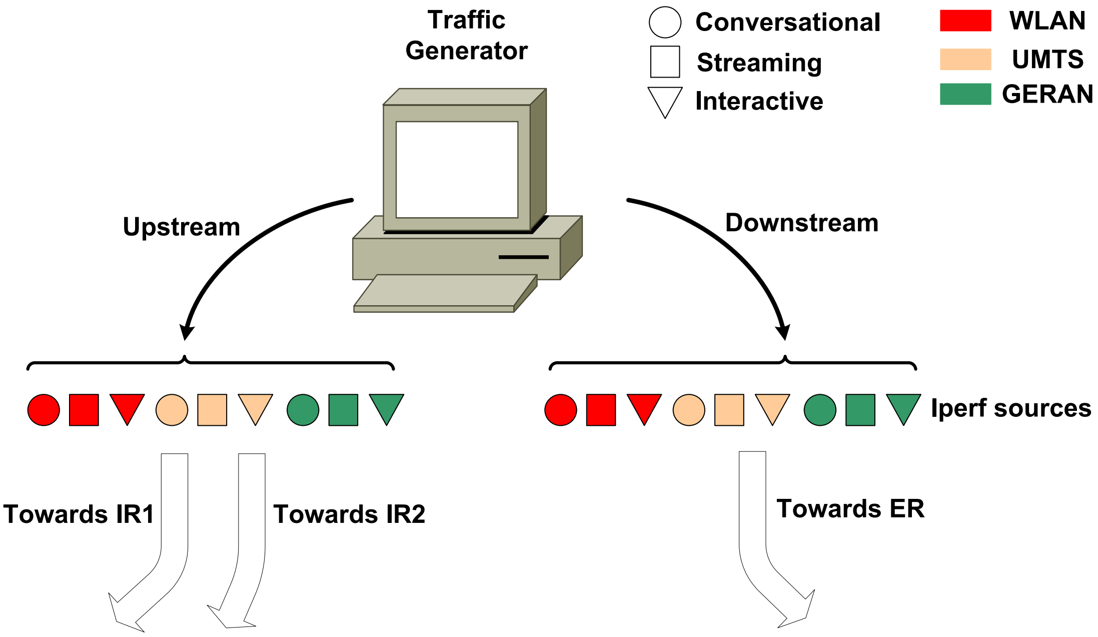


Figure 14. Coordinated traffic generation model in the CN.

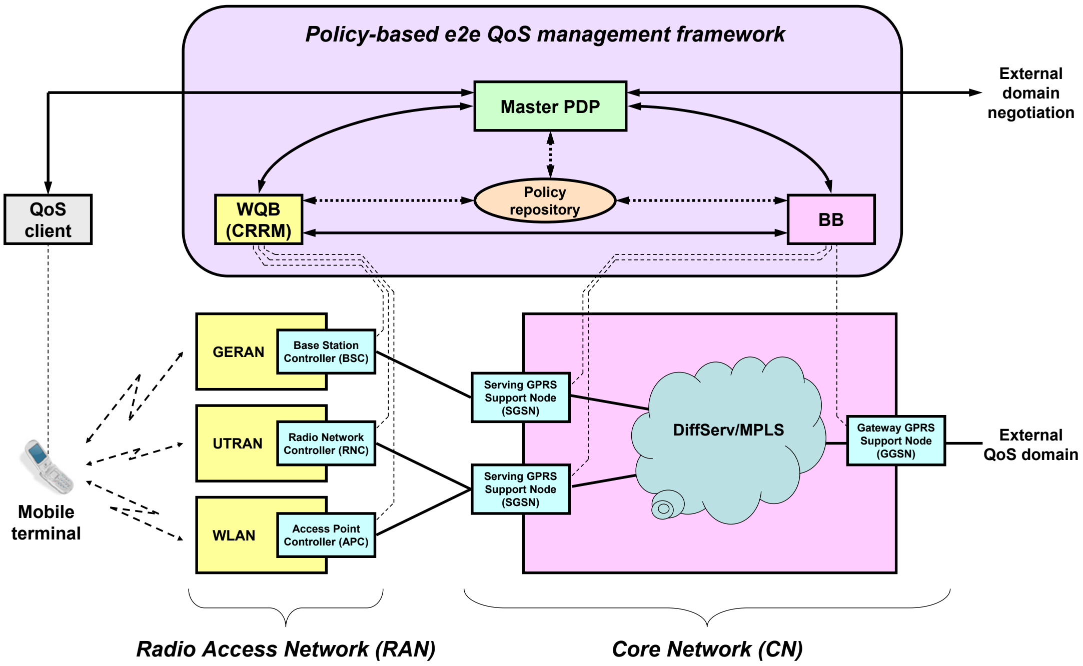

Policy-based e2e QoS management framework
Master PDP
External domain negotiation
Policy repository
WQB (CRRM)
BB
QoS client
GERAN
Base Station Controller (BSC)
UTRAN
Radio Network Controller (RNC)
WLAN
Access Point Controller (APC)
Serving GPRS Support Node (SGSN)
Serving GPRS Support Node (SGSN)
DiffServ/MPLS
Gateway GPRS Support Node (GGSN)
External QoS domain
Mobile terminal
Radio Access Network (RAN)
Core Network (CN)

Figure 15. Policy-based e2e QoS management framework.

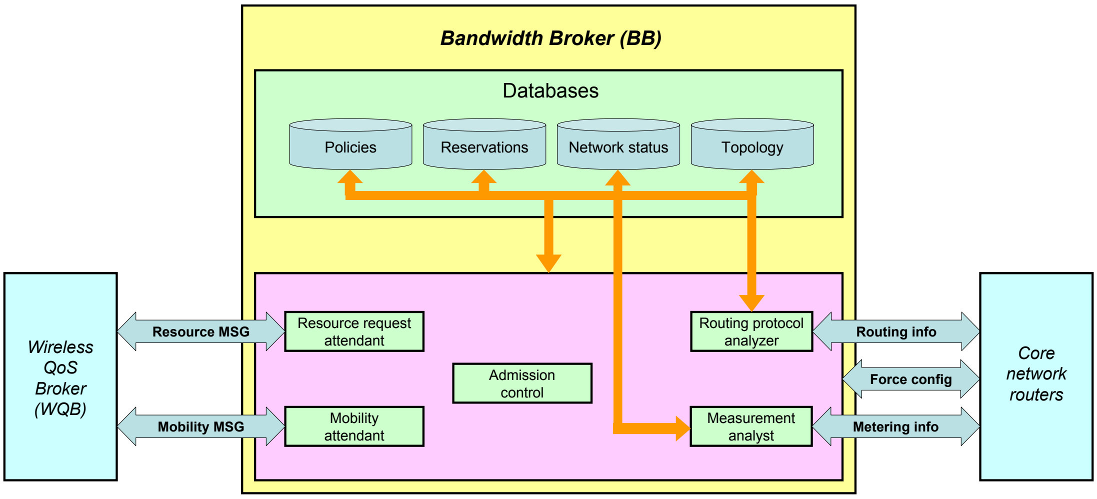


Figure 16. BB internal structure.

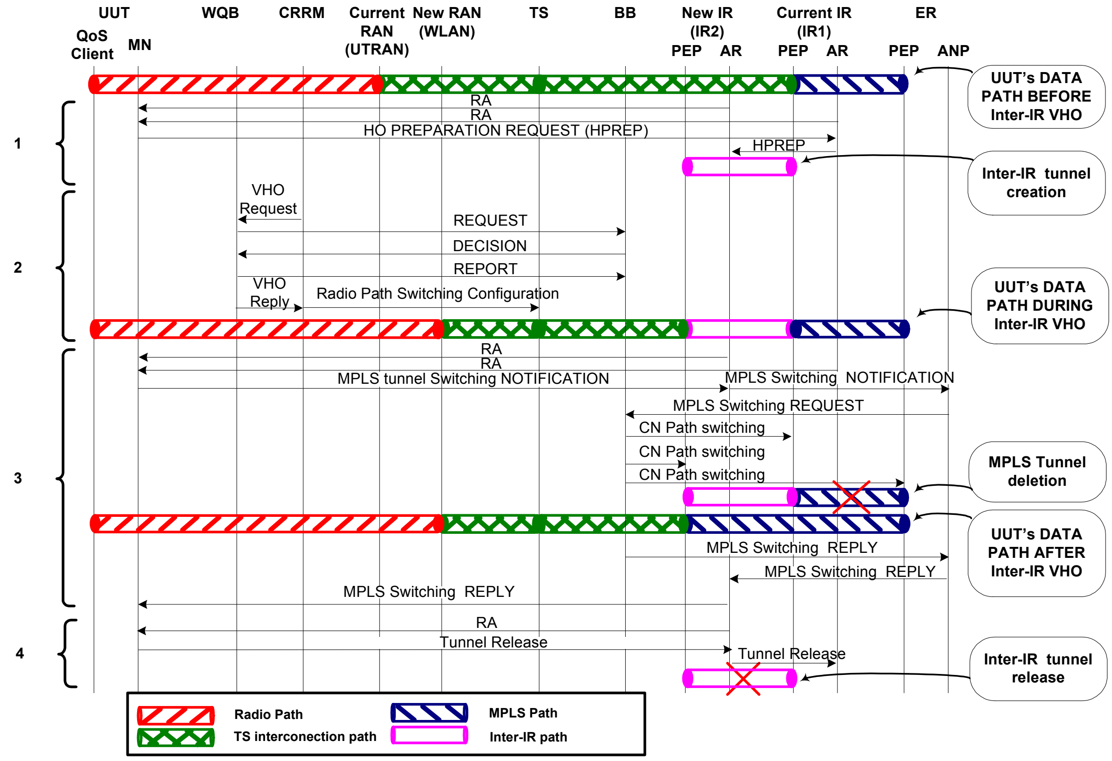
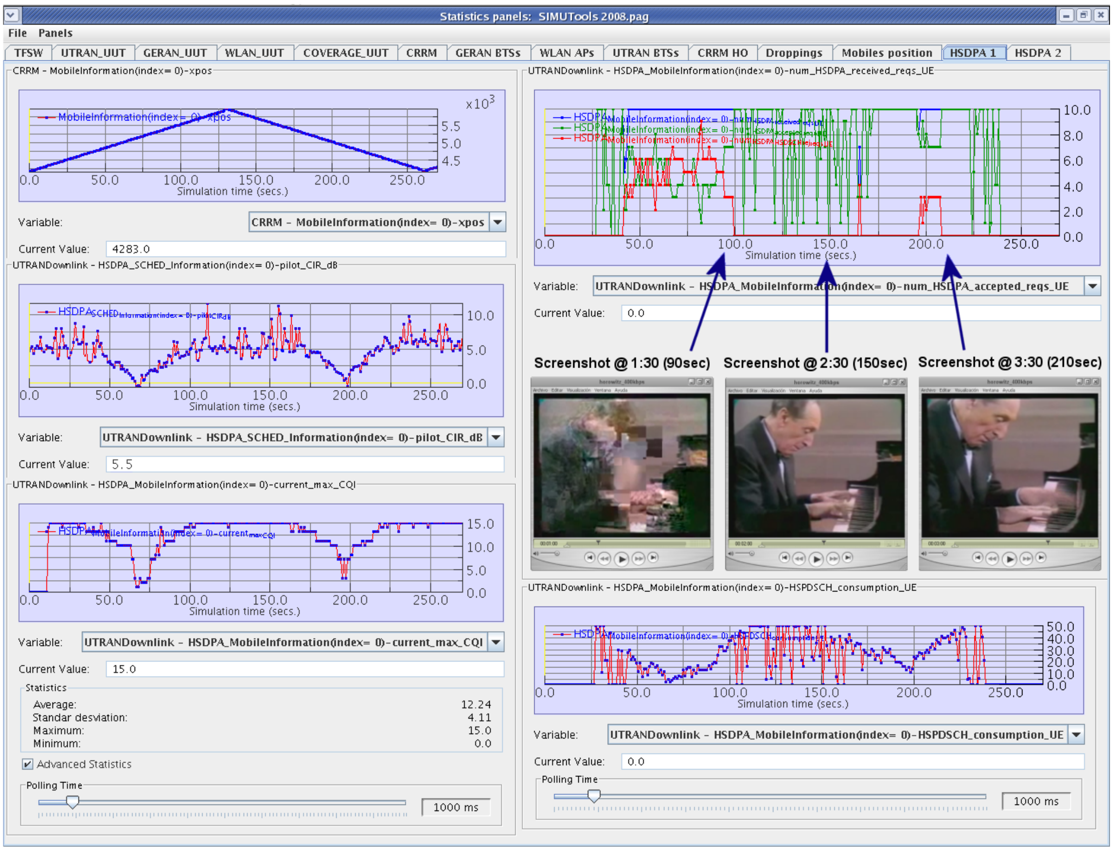


Figure 17. Signaling during inter-IR VHO with HO preparation.



CRRM - MobileInformation(index= 0)-xpos

Variable: CRRM - MobileInformation(index= 0)-xpos

Current Value: 4283.0

UTRANDownlink - HSDPA_SCHED_Information(index= 0)-pilot_CIR_dB

Variable: UTRANDownlink - HSDPA_SCHED_Information(index= 0)-pilot_CIR_dB

Current Value: 5.5

UTRANDownlink - HSDPA_MobileInformation(index= 0)-current_max_CQI

Variable: UTRANDownlink - HSDPA_MobileInformation(index= 0)-current_max_CQI

Current Value: 15.0

Statistics

Average:	12.24
Standar desviation:	4.11
Maximum:	15.0
Minimum:	0.0

☑ Advanced Statistics

Polling Time 1000 ms

UTRANDownlink - HSDPA_MobileInformation(index= 0)-num_HSDPA_received_reqs_UE

Variable: UTRANDownlink - HSDPA_MobileInformation(index= 0)-num_HSDPA_accepted_reqs_UE

Current Value: 0.0

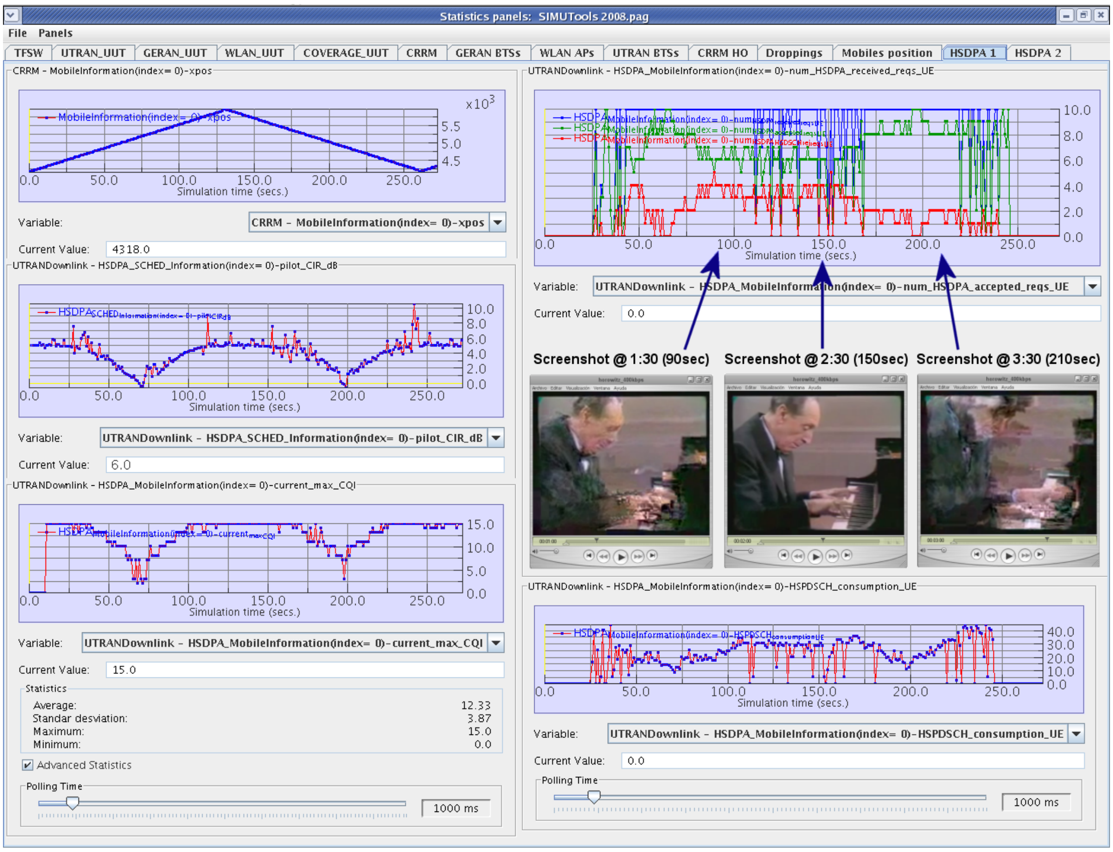
UTRANDownlink - HSDPA_MobileInformation(index= 0)-HSPDSCH_consumption_UE

Variable: UTRANDownlink - HSDPA_MobileInformation(index= 0)-HSPDSCH_consumption_UE

Current Value: 0.0

Polling Time 1000 ms

Figure 18. Real-time performance of the Maximum C/I scheduling criterion.



CRRM - MobileInformation(index= 0)-xpos

Variable: CRRM - MobileInformation(index= 0)-xpos

Current Value: 4318.0

UTRANDownlink - HSDPA_SCHED_Information(index= 0)-pilot_CIR_dB

Variable: UTRANDownlink - HSDPA_SCHED_Information(index= 0)-pilot_CIR_dB

Current Value: 6.0

UTRANDownlink - HSDPA_MobileInformation(index= 0)-current_max_CQI

Variable: UTRANDownlink - HSDPA_MobileInformation(index= 0)-current_max_CQI

Current Value: 15.0

Statistics

Average:	12.33
Standar desviation:	3.87
Maximum:	15.0
Minimum:	0.0

☑ Advanced Statistics

Polling Time 1000 ms

UTRANDownlink - HSDPA_MobileInformation(index= 0)-num_HSDPA_received_reqs_UE

Variable: UTRANDownlink - HSDPA_MobileInformation(index= 0)-num_HSDPA_accepted_reqs_UE

Current Value: 0.0

UTRANDownlink - HSDPA_MobileInformation(index= 0)-HSPDSCH_consumption_UE

Variable: UTRANDownlink - HSDPA_MobileInformation(index= 0)-HSPDSCH_consumption_UE

Current Value: 0.0

Polling Time 1000 ms

Figure 19. Real-time performance of the Round Robin scheduling criterion.

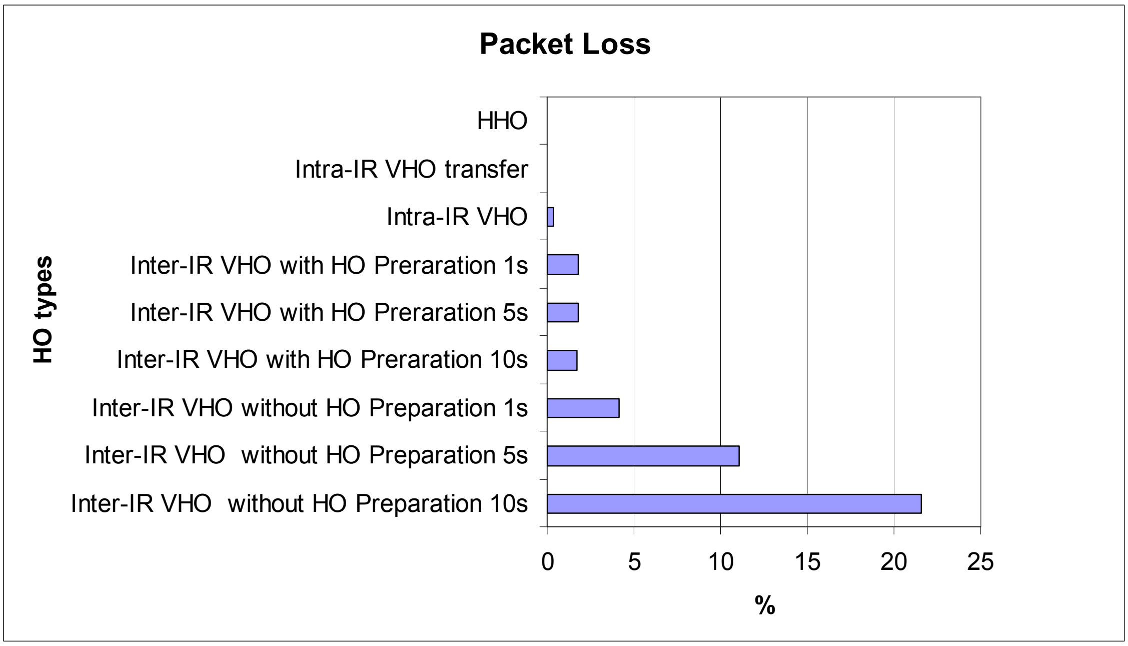


Figure 20. Average packet loss for different HO types.

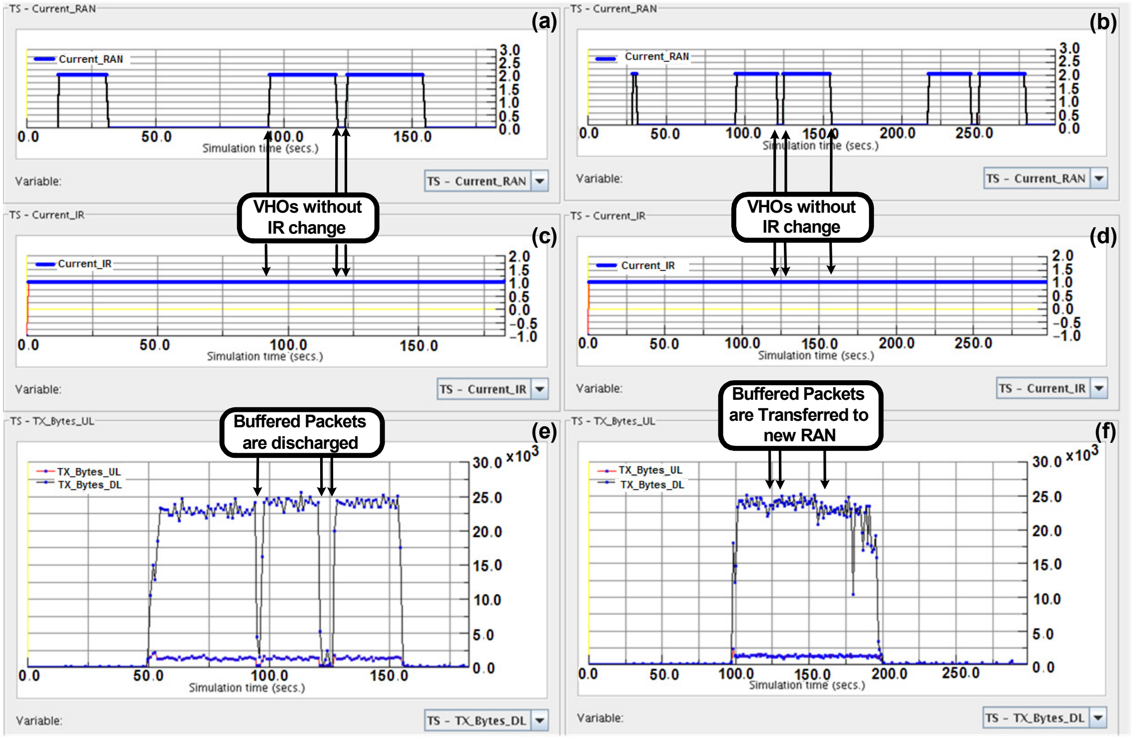


Figure 21. Intra-IR VHO without transfer policy (a), (c) and (e), and with transfer policy (b), (d) and (f).

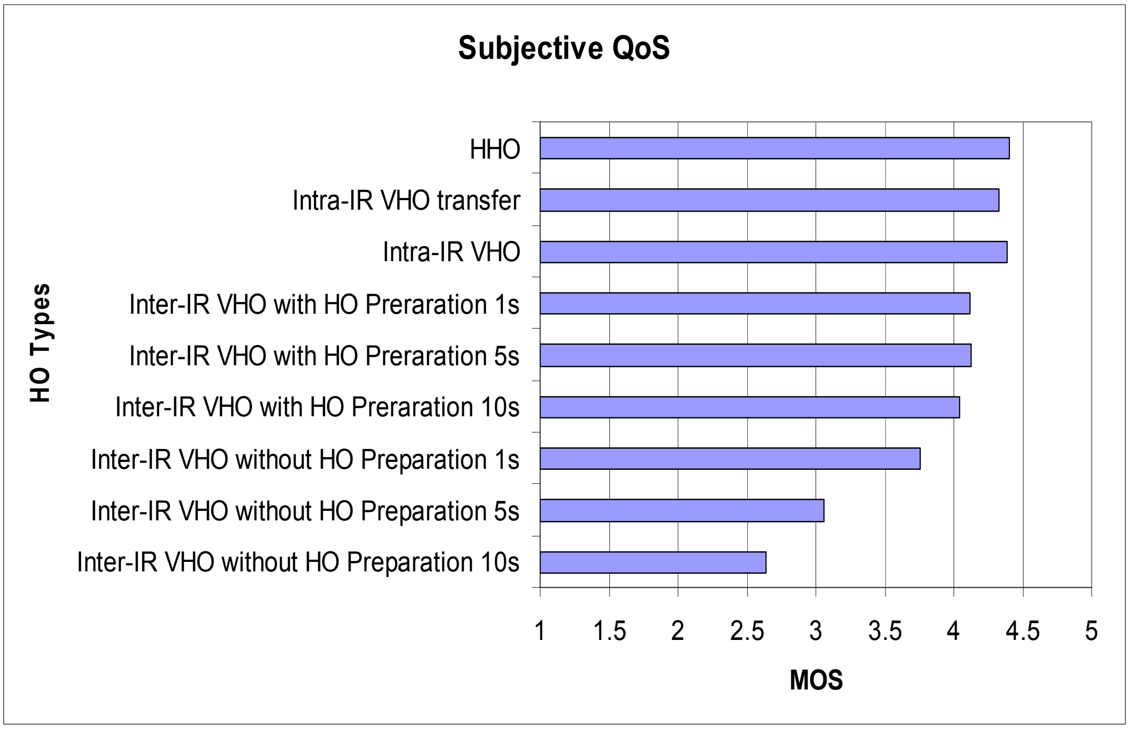
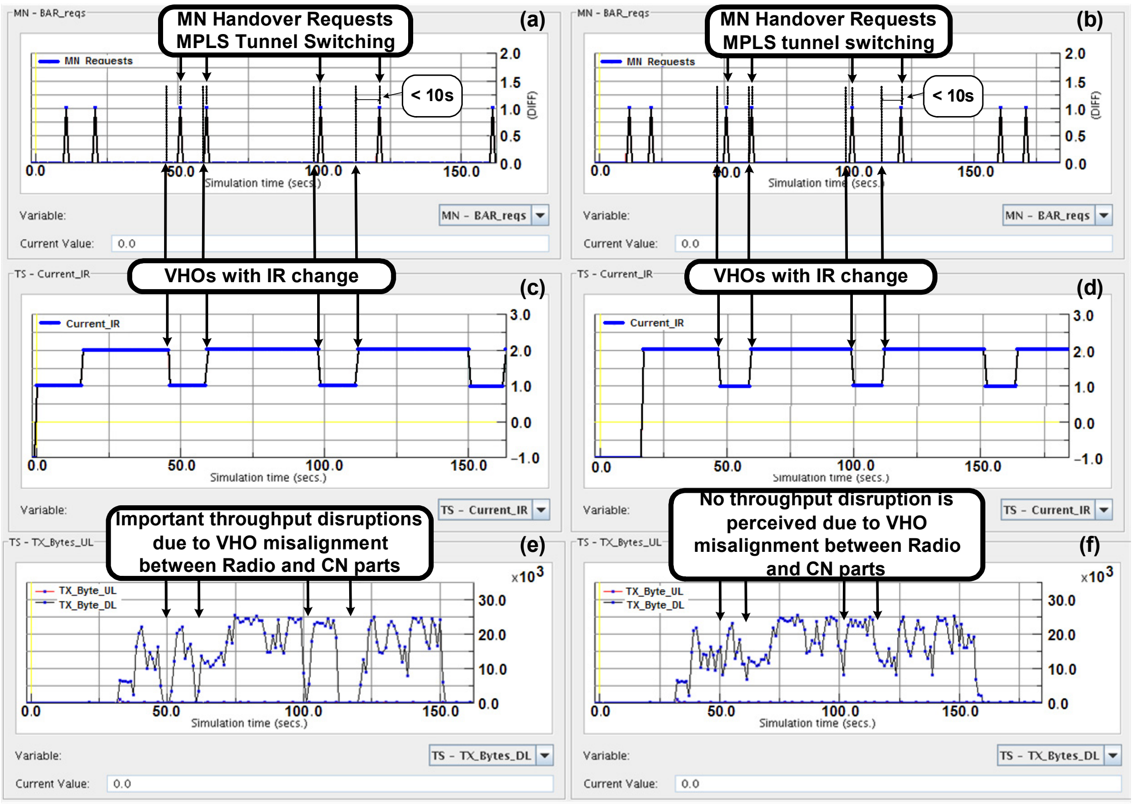


Figure 22. Mean opinion score for different HO types.



MN Handover Requests MPLS Tunnel Switching

MN Handover Requests MPLS tunnel switching

Variable: MN - BAR_reqs

Current Value: 0.0

Variable: MN - BAR_reqs

Current Value: 0.0

VHOs with IR change

VHOs with IR change

Variable: TS - Current_IR

Variable: TS - Current_IR

Important throughput disruptions due to VHO misalignment between Radio and CN parts

No throughput disruption is perceived due to VHO misalignment between Radio and CN parts

Variable: TS - TX_Bytes_DL

Current Value: 0.0

Variable: TS - TX_Bytes_DL

Current Value: 0.0

Figure 23. Inter-IR VHO with 10s RA period: without HO preparation (a), (c) and (e), and with HO preparation (b), (d) and (f).

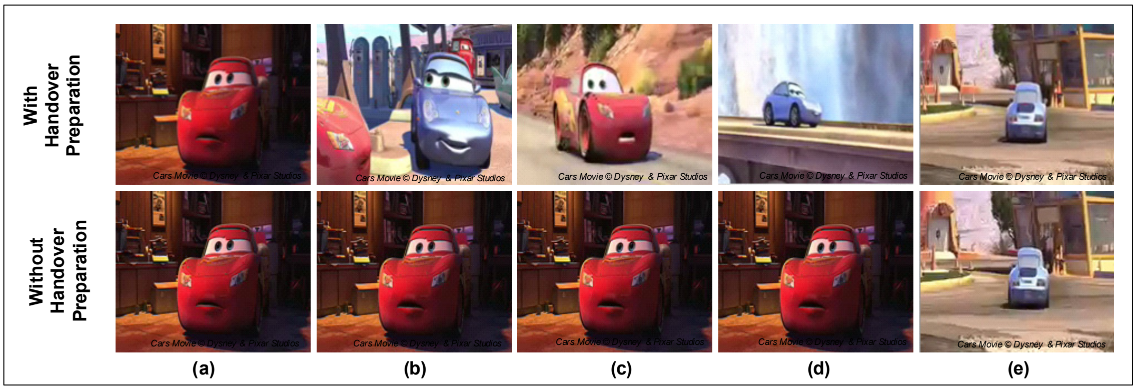

With Handover Preparation
Without Handover Preparation
Cars Movie © Dysney & Pixar Studios
(a)
(b)
(c)
(d)
(e)

Figure 24. Example of the user QoE with and without HO preparation: (a) before the VHO is triggered, (b), (c) and (d) during the VHO execution, and (e) after the VHO is completed.

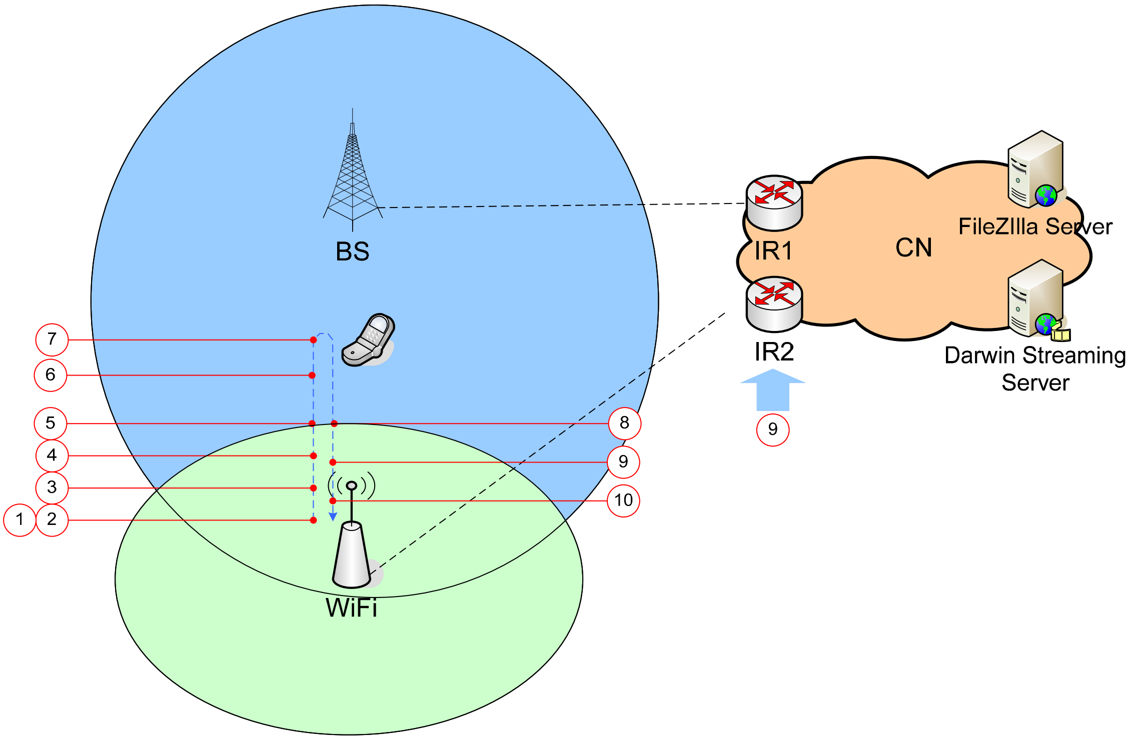


Figure 25. Test scenario for the case study.

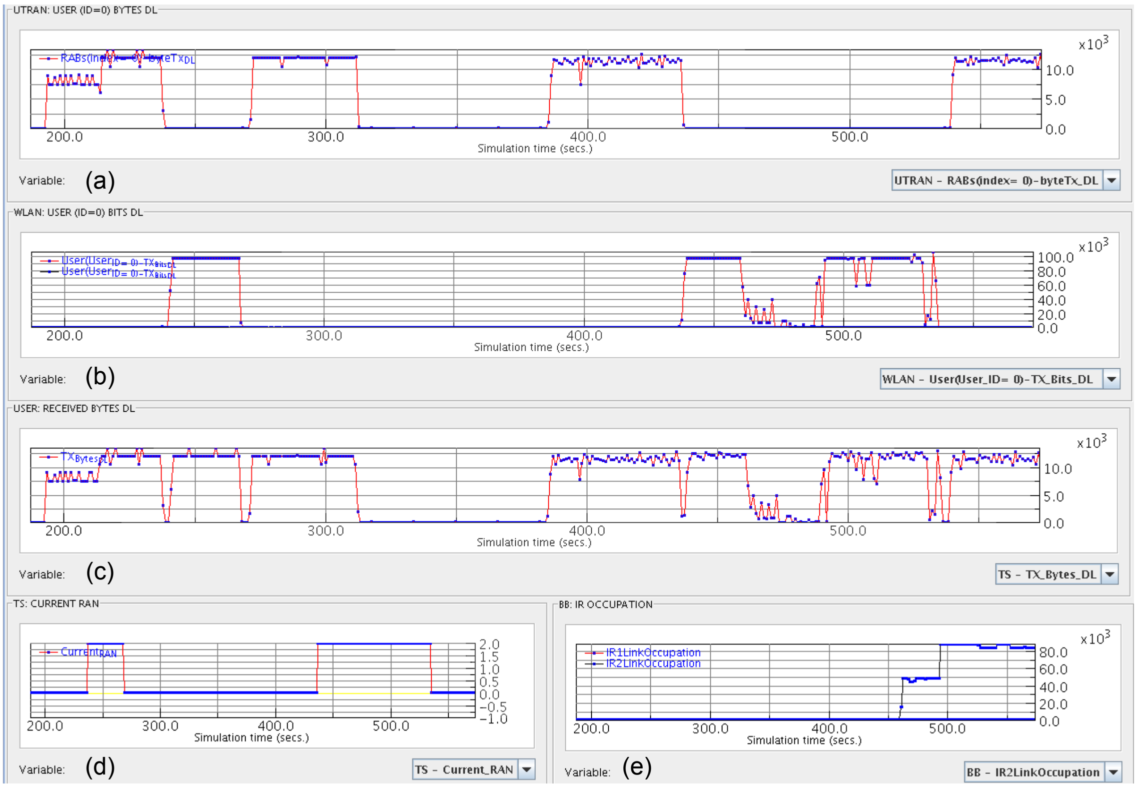
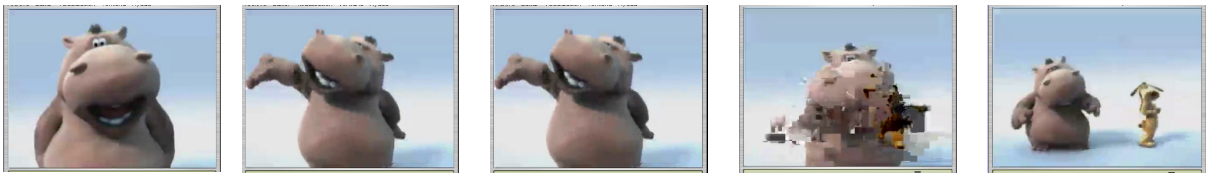


Figure 26. Real-time statistics for the case study: (a) UUT downlink throughput in UTRAN, (b) UUT downlink throughput in WLAN, (c) UUT downlink throughput at user terminal, (d) UUT current RAN, and (e) overall throughput at ingress routers.



(a) (b) (c) (d) (e)

Figure 27. Video snapshot: (a) 450s (before congestion), (b) 470s (congestion beginning), (c) 480s (frozen image), (d) 520s (distortion), and (e) 550s (recovered image).