End-to-edge QoS across Heterogeneous Wireless and Wired Domains

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

Dept. of Signal Theory and Communications
Universitat Politècnica de Catalunya
C/ Jordi Girona 1-3, Barcelona, 08034, Spain

[vucevic, fbernardo, annau, miguel.lopez]@tsc.upc.edu

ABSTRACT

Oriented to heterogeneity in every sense, next generation wireless networks are aimed to fulfil constantly increasing users' prospects. Relying on the all-IP paradigm, expectations suppose ubiquitous wireless access in heterogeneous networks. When efforts are made to make such an aspiration a reality, the main challenging issue raised is the need to seamlessly provide the user with an adequate Quality of Service (QoS) level. This paper gives detailed description of end-to-edge QoS approach in a heterogeneous wireless access network adopted in AROMA project and addresses the most important issues and observations that occurred during the implementation of such a solution in a real-time testbed. The paper gives some qualitative results on possibilities to implement coordinated interdomain QoS management by QoS mechanisms independently developed for radio and core network domains. Additionally, it gives an insight into real-time possibilities to track QoS provision across multiple domains revealing real-time behaviour before, during and after the system reconfiguration.

Keywords

all-IP, end-to-edge, heterogeneous wireless access, quality of service, real-time, testbed.

1. INTRODUCTION

Satisfaction of constantly increasing user demands in next generation wireless networks are to be facilitated by merging of technologies. Expecting ubiquitous wireless access for users in heterogeneous networks, 3GPP introduces the all-IP paradigm [1]. Hence, providing an adequate Quality of Service (QoS) level seamlessly for users in such a heterogeneous wireless/wired environment is receiving great attention and research efforts nowadays.

A mobile user is treated by Common Radio Resource Management (CRRM) [2] algorithms in order to be provided with sufficient radio resources wisely in a heterogeneous scenario. However, with the increasing heterogeneity in QoS classes, technologies and services, occasional resource scarcity in mobile backhaul may also become reality [3][4]. Thus, as part of the QoS provisioning, users' needs must be treated on an end-to-edge (e2e) level not only at session initialization, but also during entire session lifetime.

Traditionally, a wireless user is considered to experience QoS degradation due to the more restrictive radio resources. However, as the Core network (CN) is not recommended to be overprovisioned [1], the heterogeneity in radio part controlled by the CRRM algorithms should be extended to support reconfiguration in accordance with the common CN as a QoS domain. While some approaches start from developing a solution that considers a backhaul as a weighting factor in a CRRM algorithms [5], the approach we follow considers common CN and radio access network (RAN) domains as independent and relies on intercommunication between entities in charge of local QoS in the aforementioned domains. This approach is in line with ideology from [6] that is envisaged for future IP networks where CRRM is seen as the counterpart of Bandwidth Broker (BB) for the radio domain.

This paper presents an implementation of an e2e QoS approach in a real-time testbed, part of the AROMA project [7], for heterogeneous wireless access network relying on all-IP domain. The inter-domain QoS is achieved through the QoS negotiation between QoS interface of the wireless user and network resource management. The radio interface quality is controlled by a Wireless Quality Broker (WQB) designed as extension to CRRM to provide e2e QoS awareness, whereas the quality in CN is managed by a BB. In the AROMA testbed approach a master QoS decision point has been assigned to WQB.

The rest of this paper is organized as follows. Section 2 introduces the AROMA testbed and QoS baselines. Paper then focuses on the policies and rules that define the negotiation of the QoS during the entire session lifetime. Some qualitative results, giving an insight into the advantages of e2e awareness, are exposed in section 4, followed by a case study that reveals real-time application's behaviour during QoS renegotiation. The conclusions are given in section 5.

2. EVALUATION PLATFORM

The AROMA testbed is an elaborated tool developed for tests and evaluations that has been evolving through European projects. Due to the high level of complexity and implementation details of the AROMA testbed, this section focuses only on the general description of the main features of the developed tool. For more details on testbed architecture and functionalities reader is referred to [8].

2.1 General description

The AROMA testbed consists of twenty off-the-shelf Personal Computers (PCs) that allow the real-time emulation of an all-IP heterogeneous wireless access network that includes the UMTS Terrestrial Radio Access Network (UTRAN), GSM/EDGE Radio Access Network (GERAN), and Wireless Local Area Network (WLAN). The UTRAN is a release 6 version supporting High Speed Downlink/Uplink Packet Access (HSDPA/HSUPA) [9]. The three RANs rely on the common CN based on Diff-Serv technology [10] and Multi-Protocol Label Switching (MPLS) [11].

A home-made software named Communications Manager (CM) [12] deals with the issues related to the real-time execution and management. It constitutes middleware that enables synchronised communication between the testbed modules.

The administration of the testbed is achieved by means of a graphical management and configuration tool called Advanced Graphical Management Tool (AGMT). This tool has been developed to configure the initialization parameters, to control the execution, to collect logged data and present statistics in real-time [8].

For a certain user in a scenario (User Under Test - UUT) real applications may be executed. The rest of the users in the scenario are emulated in order to achieve specific load conditions where the aspects to be analyzed can be properly evaluated. This emulation has been implemented e2e, to achieve coordinate effects in radio and IP core network parts [8].

The main purpose of the AROMA testbed is to evaluate and analyse the QoS algorithms in real-time; and the effects they produce on the behaviour of real, widely available, applications in heterogeneous wireless all-IP scenarios.

2.2 QoS Framework

In this section the CRRM and BB entities are explained, as they are included in the QoS negotiation. In the AROMA testbed specifically a final decision on QoS is assigned to WQB which is detailed in next section. During Vertical Handovers (VHOs), handovers between different Radio Access Technologies (RATs), mobility management is important for QoS preservation [13]. However, the modules dealing with mobility are not directly participating in QoS negotiation and due to the limited space are not described in this paper.

2.2.1 Common Radio Resource Management

The QoS performance in a radio interface is dependent on the available RATs and the capacity they are offering. The CRRM algorithms have as a key objective the selection of appropriate RAT for an incoming user requesting service. The problem of RAT selection includes both the initial RAT selection and the VHOs. Currently the testbed incorporates six different algorithms. The two most interesting in context of future heterogeneous scenarios are: Network-Controlled Cell-Breathing (NCCB) [14] and Fittingness Factor [15].

The NCCB algorithm is addressed to heterogeneous scenarios where CDMA-based RANs (e.g., UTRAN) coexist with FDMA/TDMA-based systems (e.g., GERAN). The main idea of a NCCB algorithm, as presented in [14], is to take the advantage of the coverage overlap that several RATs may provide in a certain service area in order to improve the overall interference pattern generated in the scenario for the CDMA-based systems and, consequently, to improve the capacity of the overall heterogeneous scenario. For example, during the initial admission, the RAT selection decision is taken according to the path loss measurements in the best UTRAN cell (PL_{UTRAN}), provided by the terminal in the establishment phase. If the PL_{UTRAN} is below the path loss threshold value (PL_{th}) the user is admitted to the UTRAN, otherwise it is admitted to GERAN.

The second of the here presented RAT selection algorithms is based on the so-called fittingness factor. As explained in [15], fittingness factor is a generic CRRM metric that facilitates the implementation of cell-by-cell RRM strategies by reducing signalling exchanges and aims at capturing the multidimensional heterogeneity of beyond 3G scenarios within a single metric.

Fittingness factor (Ψ) implemented in the testbed reflects two main aspects of such multidimensional heterogeneity: the capabilities of both, terminal to support a particular RAT (i.e. depending on whether terminal is single or multimode), and the RAT to support a particular type of service (e.g. videophone is not supported in 2G networks), as well as the suitability factor, indicating the match between the user requirements in terms of QoS and the capabilities offered by the RAT. Consequently, the fittingness factor for the j-th RAT to support the s-th service requested by the i-th user with a p-th customer profile ($\Psi_{i,p,s,j}$) is calculated as a product of the corresponding capability $C_{i,p,s,j}$ and suitability factor $Q_{i,p,s,j}$ as shown in (1).

$$\Psi_{i,p,s,j} = C_{i,p,s,j} \times Q_{i,p,s,j} \qquad (1)$$
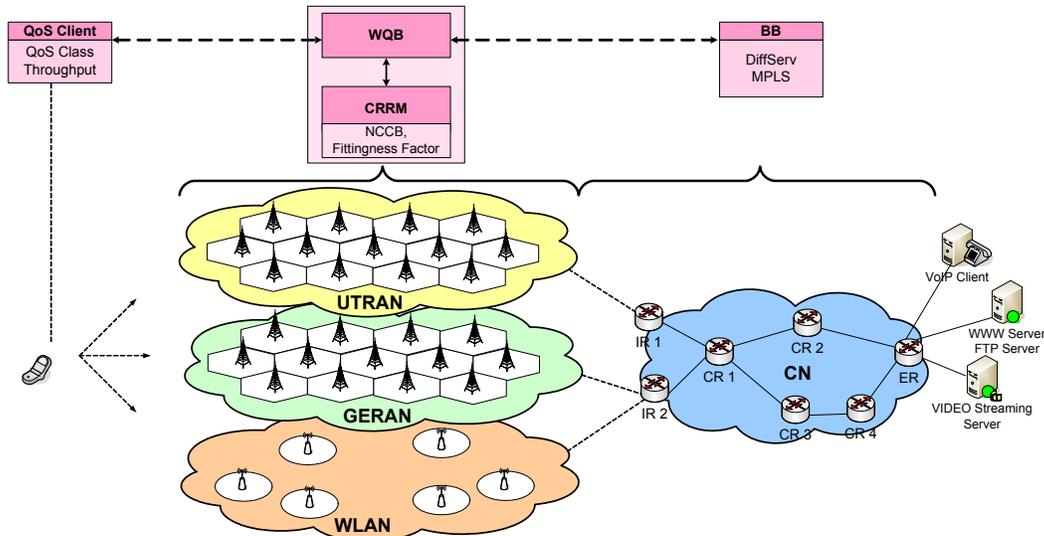
Figure 1. QoS provisioning across multiple domains in the testbed.

2.2.2 Bandwidth Broker

BB is the main architecture element of the control plane of DiffServ model proposed by IETF for supporting QoS in IP-based networks. As mentioned before, the CN is composed of DiffServ/MPLS domain with the real traffic. BB's responsibility is to control the creation and release of MPLS tunnels, while having differentiation of traffic classes, and in order to manage QoS (class transition follows previously agreed 3GPP-DiffServ class mapping [6]). Additionally, to support e2e framework, BB in the testbed is capable of suggesting the ingress router (IR) (based on the current capabilities of each attachment point) in negotiation during the session establishment, during VHOs and in case of congestion.

3. RECONFIGURATION ARCHITECTURE

In AROMA testbed, the inter-domain QoS signalling is carried out by means of a proprietary interface that finds its roots in COPS-SLS [16] framework. COPS is a hierarchical client-server protocol that defines a PDP (Policy Decision Point) and a PEP (Policy Enforcement Point). Thus, AROMA e2e QoS signalling is based on a three handshake protocol between entities that enables the exchange of QoS parameters and decisions.

In the AROMA testbed, a master PDP functionality has been assigned to WQB. Its implementation makes it closer to radio interface in the testbed, which is a consequence of the fact that the UUT is wireless. Therefore, WQB may be seen as an extension to basic CRRM entity, and is physically part of the same module. However, it is performing the decision making process from a high layer point of view where a complete vision on the network status is known. The functionalities of WQB as being independent of the CRRM calculations make it virtually a separate module.

The functional representation of the e2e QoS architecture in AROMA testbed is shown in Figure 1. The three radio domains managed by CRRM include WQB as an extension that facilitates communication with QoS Client (user's interface for QoS specification) and BB. BB is in charge of the CN composed of seven routers: two IR, one egress router (ER) and four core routers (CR).

The presented architecture supposes tables containing QoS specification of the user. In the example these are throughput and QoS class, but may be easily extended to more complex demands. The table is created upon session establishment and later on updated in real-time with information on the RATs that are feasible for that user and the IRs (CN paths) that these RANs are connected to.

3.1 The Negotiation Procedures

The e2e negotiation is recognized in three forms: initiated by the user (QoS Client), initiated by CRRM or initiated by BB. All the mentioned procedures during negotiation are illustrated in Figure 2. The given examples cover all the possible general cases, where dashed lines mark negotiation that optionally appears depending on the negotiation status as the entities are contacted only when necessary. The communication between WQB and QoS Client and BB is the aforementioned three handshake procedure including request (REQ), decision (DEC) and report (RPT). WQB and CRRM as closely related modules exchange simple demand (DEM) – confirm (CONF) message pairs.

User initiates negotiation when connecting/disconnecting from the system in order to achieve the desired QoS, as well as when modifying QoS preferences (class, throughput, etc.) during an active session.

Whenever the CRRM decides to change the preferences for the user (due to the loss of coverage or change in network
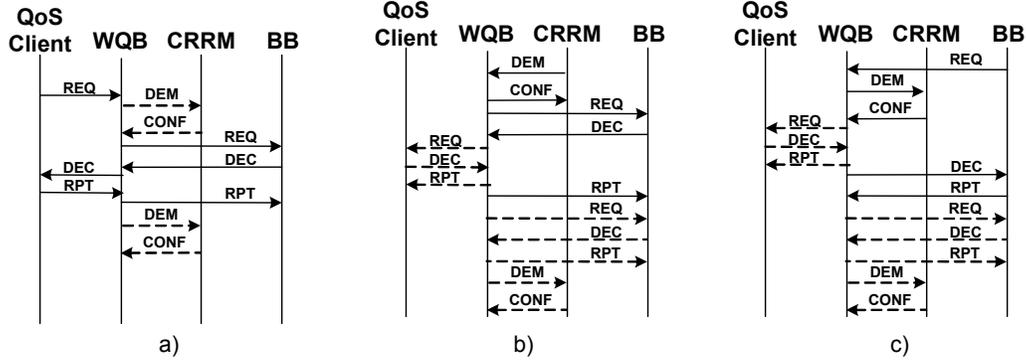
Figure 2. The e2e re-negotiation procedures initiated by a) the user, b) CRRM, c) BB.

preferences), the resulting action is VHO approval, VHO disapproval with session preservation or VHO disapproval with session dropping.

The additional possibility in AROMA testbed is to have BB that starts re-negotiation in order to inform the WQB on changes in state of the CN that may lead to QoS degradation (even suggest VHO if feasible).

The implementation of the negotiation also assures that whenever one of the negotiation procedures is in execution, any new negotiation regarding the same user is ignored.

3.2 Decision Making

The WQB takes the final decision based on the parameters passed from the CRRM and information received from BB. It is worth mentioning here some observations regarding decision exchange.

On one side, the preferences CRRM delivers to WQB depend on the employed RAT selection algorithm. As these are two virtually separate modules, the suitability of the different CRRM RAT selection algorithms to e2e negotiation may not be the same. The information CRRM provides to WQB is a table of candidate RATs for that user and the suitability of each one of them. In case that several RATs are suitable certain preferences must be extractable, usually expressed through resulting weights. In the case of NCCB algorithm, the nature of the algorithm itself bases on measurements on one RAT (UTRAN) and is assigning users to it or to GERAN accordingly (see [13]). This way the NCCB algorithm gives only one RAT as the possible solution for a user, disabling the possible change of preference by WQB through further communication with the CN.

On the other side, it is worth mentioning that the preferable behaviour is to have flexible QoS in the CN in the sense the BB does not reject a session even if the resources are scarce on that IR. The BB provides the WQB with QoS characteristics of interest (throughput, delay, loss, etc.) and lets WQB extract the preferences. As user's throughput is usually quite negligible in comparison to overall IR capacity, this is supposed to prevent connection loss with possi-

bility to stay connected with lowered QoS guarantees. In that case QoS Client is informed on QoS degradation, and may reject connection himself later on.

When interdomain QoS negotiation is enabled another occurrence that should be considered with caution is the way the user is treated after the VHO execution. Namely, CRRM algorithms usually have hysteresis that protects a user from ping-pong effect in cellular networks. Similarly, after changing current RAN upon a CN suggestion, the terminal is also preserved from immediately being submitted for another VHO due to CRRM preferences by putting a back-off period, either in CRRM or WQB, that prevents VHO if not strictly necessary.

4. CASE STUDIES

The AROMA testbed enables vast variety of tests that may be preformed in real-time. In this section some representative case studies that regard the issues of e2e QoS awareness are presented.

4.1 Extended QoS awareness in radio

The suitability of the CRRM algorithms to the presented e2e QoS solution is discussed in this case study. The evaluation consists in four scenarios under two CN conditions and with two different CRRM algorithms.

In all the tests, UUT moves diagonally through the area under study ($4x8$ km^2) covered with 13 UTRAN and 13 GERAN co-located base stations, and 6 WLAN hotspots (as in Figure 1). The UUT is running conversational service in test mode and is connecting/disconnecting from the system every 30s. The resulting statistics are based on 250 consecutive connections in each test.

Four scenarios were evaluated to give diversity in network load: scenario 1 (S1) with 300, scenario 2 (S2) with 600, scenario 3 (S3) with 900, scenario 4 (S4) with 1200 conversational users. Average session length of emulated users is 180s, with arrival rate of 0.8 arrivals/hour/user.

The employed CRRM algorithms in tests were fittingness factor and NCCB. All the tests were repeated for the case

GERAN Selection
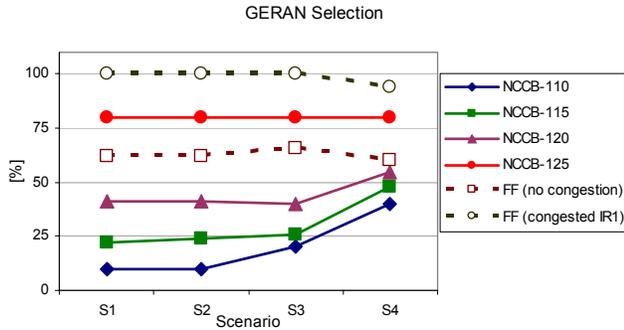
Figure 3. Probability to assign GERAN to voice user in different network conditions with different CRRM policy setup.

with no congestion in CN and in case when IR1 (UTRAN's access point) is congested with additional traffic. In the second case, BB is aware that the QoS may not be supported in UTRAN due to the congestion.

In Figure 3 dashed lines present the RAT assignments for the terminal when fittingness factor algorithm is used (denoted with FF). The fittingness factor parameters are set so the session assignment in ~66% of the times was to GERAN and in rest of the times to UTRAN, for the case with no CN congestion. When IR1 was congested e2e negotiation made terminal connected to GERAN in 100% of the cases for the first three scenarios, and in some 92% in scenario 4. In the scenario 4 the 8% of the times terminal was connected to UTRAN due to the congested radio interface.

When tests were repeated for NCCB (solid lines in Figure 3) no difference occurred in RAT assignment when IR1 was congested or not congested. As explained before, the NCCB algorithm decides on the RAT selection based on UTRAN path loss measurements. The Figure 3 shows the RAN assignment for a tested user in four cases, when the threshold value is changing from 110 to 125 dB for all the scenarios. This demonstrates that the distribution of the conversational sessions between GERAN and UTRAN depends on the scenario (network load) and on CRRM policy parameter. However, even when one of the IRs is congested the results on RAT selection remain unchanged when e2e negotiation is applied over NCCB.

It is worth noting that the aim of this case study is not to enter in details with parameter adjustment, but rather to show how the e2e awareness in session connecting impacts on the final RAT selection decisions. The obtained results demonstrate that the finally selected RAT may differ significantly when both RAN and CN parts influence on the decision with respect to the case in which the CRRM policy is the absolute determining factor.

4.2 QoS driven System Reconfiguration

This case study demonstrates the behaviour of the system and the user that result in system reconfiguration; and also

demonstrates how the influence on real applications may be tracked. For this case study, the model that is tested considers a real user moving through space under constant UTRAN coverage and occasional WLAN coverage, as in Figure 4 (showing only the entities from Figure 1 that are of relevance for this case study). The UUT uses two services, first FTP application to download a file from the server, followed by the video streaming service. During a session UUT and system perform a set of actions that involve QoS renegotiation.

The time instances of all the actions can be followed in Figure 4 and are explained in Table 1 for the easier tracking of real-time captured throughput statistics from Figure 5. Figure 5 gives an insight into throughput corresponding to UUT in UTRAN, WLAN, and at user itself; the current RAN the UUT is connected to (UTRAN=0, WLAN=2), which enables easier VHO tracking; and overall throughput on IRs.

The QoS renegotiation is triggered 5 times in the scenario: two times by QoS Client for the session modification (actions 3 and 4), two times by CRRM (actions 5 and 8) and once by BB (action 10). The first negotiation triggered by the user includes the communication of QoS Client with WQB, and WQB with CRRM. The BB is not contacted in that case as the CRRM modifies the session's throughput on the same RAT (UTRAN). However, when user decides to change the class, WQB also communicates the BB as the network polices in CRRM express desire to switch the RAT (resulting in VHO to WLAN).

The first reconfiguration initiated by the CRRM is due to the loss of coverage when the need for VHO appears. The second one appears when the coverage of WLAN is again detected, so the radio part preferences suggest a VHO (to WLAN). In both cases the CN is contacted to check if the reconfiguration is possible. Note that in the first case the session would be dropped if reconfiguration request is rejected by CN, while in the second one it would stay active on the same RAN.
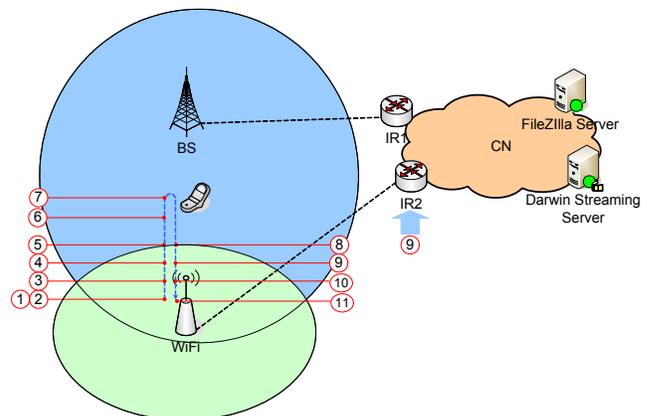
Figure 4. The test scenario, marking user's movement and action instances.

Table 1. Actions in test scenario

	Action	Description	Time instance
1	User starts interactive session	64kb/s in downlink	
2	User starts file download	FTP client/server [17]	~190s
3	User modifies session throughput	Increase bandwidth in downlink to 96kb/s	~220s
4	User modifies session class to Streaming (same throughput)	Results in VHO, due to CRRM preferences for that class	~240s
5	CRRM initiates VHO	Due to loss of coverage of the current RAT	~270s
6	User closes FTP application		~310s
7	User starts video streaming [18][19]	Video 64kb/s + Audio 24kb/s	~380s
8	CRRM initiates VHO	Due to CRRM preferences	~440s
9	Congest IR2 (current CN access point)	Additionally generated traffic [20]	~460s
10	BB suggests reconfiguration (VHO)	Results in VHO, due to the congestion detection	~540s
11	User stops session		
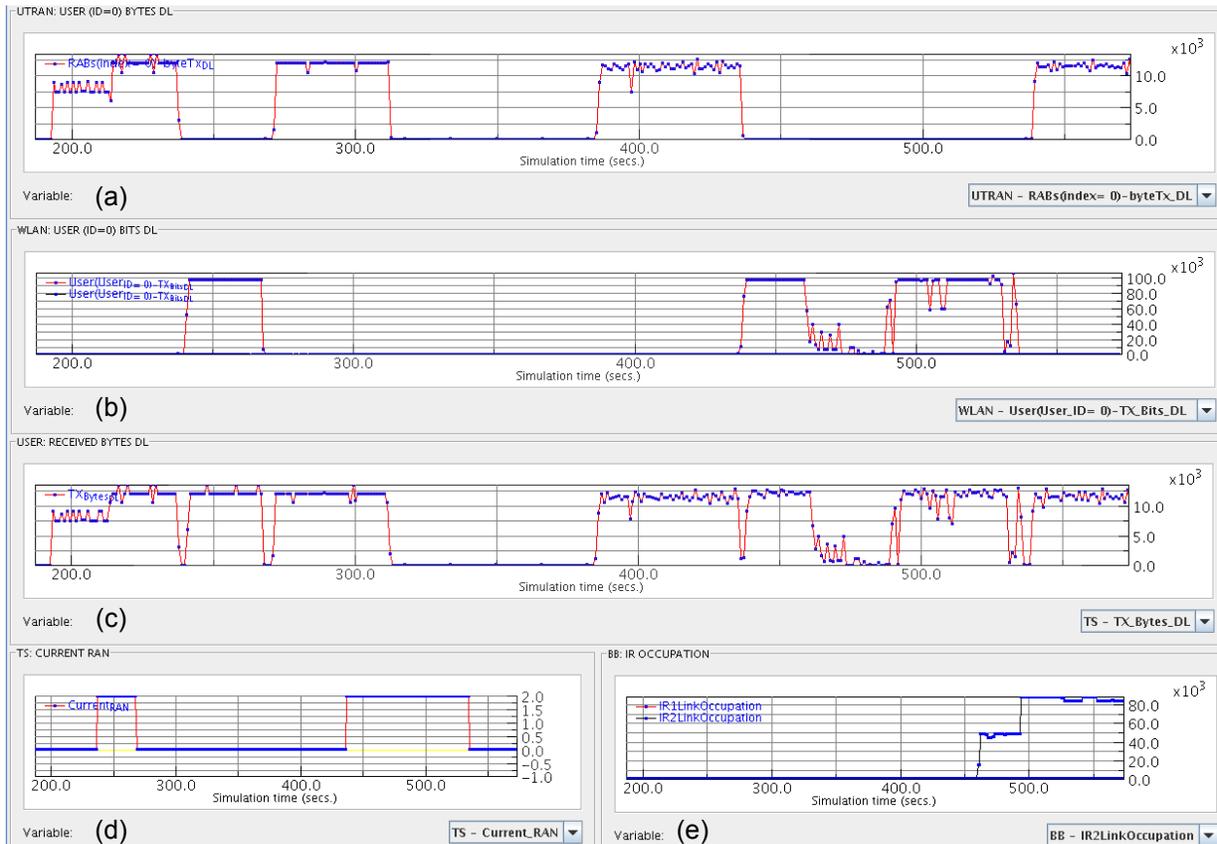
Figure 5. Snapshot of AGMT real-time statistics: downlink throughput of the UUT a) in UTRAN, b) in WLAN, c) at user terminal; d) current radio access technology UUT is connected to; e) overall throughput at ingress routers.

Figure 6. Video snapshots at: a) 450s (before congestion), b) 470s (congestion beginning), c) 480s (frozen image), d) 520s (distortion), e) 550s (recuperated image).

The reconfiguration triggered by the CN is consequence of the congesting by additional load generated at IR2 (to which WLAN is connected). This example is only a concept proving, that demonstrates the possibility that BB detects congestion and suggest session VHO (to IR1 that bares UTRAN) with aim to preserve QoS.

In all the reconfiguration cases the communication with QoS client is enabled for the case when QoS degradation is inevitable. However this is not a case of this example.

Figure 6 shows tracked video degradation of real application during CN congestion. The most common effects are frozen or distorted image. By means of testbed this applications may further be used for evaluation of objective quality perception by a user [21].

5. CONCLUSIONS

In this paper, the description of an approach to end-to-edge QoS awareness for ubiquitous wireless access is given. The procedures that enable QoS renegotiation in heterogeneous radio access environment relying on a common all-IP core are shown, disscussing some of the issues when such an approach is implemented. Especially, the coupling between RAT selection algorithms and e2e QoS management has been shown to be crucial. Quantitative behaviors have been compared for two RAT selection algorithms, Fittingness Factor and Network-Controlled Cell-Breathing, reflecting differences in flexibility to adapt to possible end-to-edge preferences. All the tests were performed over the real-time heterogeneous wireless all-IP testbed. This enabled a real-time insight into activities that justify QoS renegotiation, both on the system and on the real multimedia application. By means of real-time testbed it has been illustratively shown that the end-to-edge QoS re-negotiation helps provision of a proper QoS level to the final user, with an example in which the reconfiguration actuates during a real application session.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] 3GPP TR 22.978, *All-IP Network (AIPN) feasibility study*, June 2006, Available: http://www.3gpp.org/

[2] Tolli, A., Hakalin, P., Holma, H., "Performance evaluation of common radio resource management (CRRM)", IEEE ICC 2002, Vol. 5, May 2002, pp. 3429-3433

[3] Sagfors, M., Virkki, V., Kuningas, T., "Overload Control of Best-Effort Traffic in the UTRAN Transport Network", IEEE VTC 2006-Spring, Vol. 1, May 2006, pp. 456-460

[4] A. Bolle, H. Herbertsson, "Backhaul must make room for HSDPA", *Wireless Europe*, Issue 43, March 2006, pp.17-18

[5] Gunnar, A., Ahlgren, B., Blume, O., Burness, L., Eardley, P., Hepworth, E., Sachs, J. Surtees, A., "Access and Path Selection in Ambient Networks", in Proc. of the 16th IST Mobile and Wireless Communications Summit, July 2007, pp. 1-5

[6] Raisanen, V., Nokia Networks OY Finland, *Implementing Service Quality in IP Networks*, John Wiley & Sons, 2003

[7] IST AROMA Project, 6th Framework Program of the European Community, http://www.aroma-ist.upc.edu

[8] Umbert, A., López-Benítez, M., Bernardo, F., Vucevic, N., Azevedo, R., Oliveira, A., "The real-time AROMA testbed for all-IP heterogeneous wireless access networks", ICST TRIDENTCOM 2008, March 2008

[9] López-Benítez, M., Bernardo, F., Vucevic, N., Umbert, A., "Real-time HSPA emulator for end-to-edge QoS evaluation in all-IP beyond 3G heterogeneous wireless networks", ICST QoSim 2008, march 2008

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

[11] Rosen, E., Viswanathan, A., Callon, R., *Multiprotocol Label Switching Architecture*, IETF RFC 3031, January 2001

[12] Revés, X., Umbert, A., Ferrús, R., Cassadevall, F., "Implementation of a Real-time UMTS Testbed in a PC Network", in Proceedings of ACIS 3rd SNPD, June 2002, pp. 345-352

[13] Bernardo, F., Vucevic, N., Umbert, A., López-Benítez, M., "Quality of experience evaluation under QoS-aware mobility mechanisms in the AROMA testbed", European Wireless Conference 2008, June 2008

[14] Pérez-Romero, J., Sallent, O., Agustí, R., Wang, L., Aghvami, H.,"A Novel Algorithm for Radio Access Technology Selection in Heterogeneous B3G networks", IEEE VTC 2006-Spring, May 2006, pp. 471-475

[15] Pérez-Romero, J., Sallent, O., Agustí, R., "A Novel Metric for Context-Aware RAT Selection in Wireless Multiple Access Systems", The IEEE ICC 2007, June 2007, pp. 5622-5627

[16] Thi Mai Trang Nguyen, Boukhatem, N., Doudane, Y.G., Pujolle, G., "COPS-SLS: a service level negotiation protocol for the Internet", *IEEE Communications Magazine*, Vol. 40, Issue 5O, May 2002, pp. 158-165

[17] FileZilla Client/Server. FileZilla Project, available at: http://filezilla-project.org

[18] QuickTime Pro 7, available at: http://www.apple.com/quicktime/

[19] Darwin Streaming Server, available at: http:// developer.apple.com/opensource/server/streaming/

[20] Iperf traffic generator, available at: http://dast.nlanr.nat/Project/Iperf/

[21] Vucevic, N., Bernardo, F., Umbert, A., Budzisz, L., "Evaluation of Perceived QoS with Multimedia Applications in Heterogeneous Wireless Access Networks", IEEE ISWCS 2007, October 2007, pp. 102-106