

Towards End-to-end QoS in a Beyond 3G Scenario

R. Ferrús[♦], N. Nafisi[□], O. Sallent[♦], J. Pérez-Romero[♦], A. Gelonch[♦]

[♦]Signal Theory and Communications Department, UPC, Barcelona

[□]Centre for Telecommunications Research, King's College London

e-mail: [ferrus, antoni, sallent, jorperez]@tsc.upc.edu [nima.nafisi]@kcl.ac.uk

Abstract.- This paper presents the pillars for an end-to-end QoS framework in IST-AROMA project. This includes a QoS architecture framework encompassing common radio resource management (CRRM) for the heterogeneous radio access network part and enabling QoS coordination with QoS-enabled IP-based segments, either serving as transport option within the 3GPP R6 Radio Access Network (RAN) or as main building block for an all-IP architecture aligned to 3GPP Long Term Evolution (LTE) vision. The envisaged technical approach is supported by the progress achieved within the IST FP6 EVEREST project, which is partly reflected in the paper as well.

I. Introduction

A significant challenge for beyond third generation (B3G) wireless networks is to coordinate the many different types of networks being used for wireless communications. A plethora of different radio access networks (e.g. cellular networks as GPRS and UMTS as well as wireless LAN hostspots) will have to co-exist. These different networks ought to be inter-connected in an optimum manner with the ultimate objective to provide the end-user with the requested services and corresponding QoS (Quality of Service) requirements. In particular we assume a future mobile network scenario where different RANs may co-exist in the same area and multi-mode terminals can be connected to them. Furthermore, common radio resource management (CRRM) mechanisms are expected to be deployed within the different RATs of the B3G network as well as QoS control mechanisms in IP core network. The objective of the EVEREST project was to devise and assess a set of specific strategies and algorithms for access and core networks, leading to an optimised utilisation of scarcely available radio resources for the support of mixed services with end-to-end QoS mechanisms. In order to achieve this end-to-end QoS objective, a functional architecture to manage QoS in a B3G network has been proposed and analysed. The approach chosen is centralised, based on the BB (Bandwidth Broker), as it is thought that a centralised QoS management of the BB can provide a simple interface between the IP QoS and the CRRM in the radio access network. This latter interface has been achieved by the definition of a new functional entity named Wireless QoS Broker (WQB) that behaves as the counterpart of the BB within the radio access domain.

This paper gives an overview of the key building blocks for the end-to-end QoS framework considered in IST-AROMA project. The envisaged technical approach is supported by

the progress achieved within the IST FP6 EVEREST project which is partly reflected in the paper. In the first section of the paper, the rationale of the proposed QoS framework is discussed and the main study items arisen from the proposal are highlighted. Then, sections III to VI are targeted to cover which activities and achievements have been carried out in the EVEREST project under such relevant study items. After that, section VII discusses how AROMA project is expected to leverage EVEREST's achievements so that a progressive transition towards an all-IP radio access network is pursued. Finally, last section concludes the paper.

II. Reference QoS Framework

The B3G network architecture proposed in EVEREST is based on 3GPP Release 6 network architecture. As such, a common IP packet network with diffserv QoS mechanisms is considered for the core network (CN) where different radio access networks (UTRAN, GERAN) are attached. Mobility and session management protocols in the CN are those specified in the 3GPP. WLAN access is also taken into account by assuming a tight coupled interworking. IP transport is assumed to be present from radio network controllers (i.e. RNC, BSC) up to network gateways (i.e. GGSN).

Then, the motivation for an end-to-end QoS architecture and the coordination between the IP CN and RANs can be formulated as the following. In a mobile access network with heterogeneous RANs, the UE (User Equipment) has a high probability of being in the range of several APs (Access Point, following the terminology in RFC3753) with the same or different radio access technologies. Moreover these APs might be connected to different radio network controllers and the paths from each radio network controller to one of the gateways of the mobile access network may present different IP QoS parameters in terms of jitter, bandwidth and packet loss guarantees. The usual AP selection among a set of candidate APs is done based on RRM parameters. Once this selection has been done, then the IP QoS reservation is carried out; and based on the successful output of this reservation (for instance in 3GPP the message "Activate PDP Context Accept") an end-to-edge communication path is set up from the UE to the gateway of the B3G access network. However, this approach for end-to-edge QoS and session establishment is not optimised as it considers IP QoS parameters only a posteriori. Other approaches are possible, as suggested in [1],[2],[3] and [4]. In these papers, prior to AP selection, coordination between the RRM entities of the RANs and the IP QoS management is performed. Thus, in a B3G

system where the IP access network supports different radio access technologies and can therefore become a source of congestion, a QoS-aware path selection mechanism has to be provided in order to avoid the connection through an AP without the required QoS guarantee in the access network part.

In order to achieve this end-to-end QoS objective, a functional architecture to manage QoS in a B3G network has been proposed and analysed. The QoS management framework addresses the QoS provisioning problem of multimedia services (e.g. IMS services) in B3G systems. The proposed framework is based on the UMTS QoS architecture introduced in 3GPP release 5 for IMS services [5]. Under such 3GPP framework, a Policy Decision Function (PDF) is used to authorise IP QoS parameters for multimedia services. This 3GPP QoS framework can be seen as a first step to introduce policy-based mechanism in UMTS networks. Over such a basis, several extensions have been proposed in IST-EVEREST project so as to fulfil B3G scenarios requirements. In particular, a policy-based RAT selection decision function is considered mandatory within the QoS management framework as well as the introduction of dynamic QoS negotiation mechanisms among all the potential RATs and the CN.

The proposed architecture solution over a B3G network, illustrated in Figure 1, intends to achieve the above requirements and allows for an optimised management, considering both the core network part and the heterogeneous radio access part. As shown in the figure, an IP diffserv-enabled core network provides connectivity between radio network controllers and SGSN/GGSN nodes. QoS management is this CN is achieved by a Bandwidth Broker (BB) concept [6]. The BB is in charge of the control plane of the DiffServ domain. Then, the Wireless QoS Broker (WQB) concept has been proposed in EVEREST to account for common radio resource management and QoS control in the heterogeneous radio access part. The WQB can be understood as the counterpart of the BB for the radio part of the access network [11]. Finally, the PDF entity already introduced in 3GPP R5/R6 policy framework is maintained but its functions extended in order to cope with QoS negotiation with external peer domains involved in the provisioning of end-to-end services. So, we refer to this new extension as Master PDF (MPDF) in Figure 1. Relationship between the MPDF and the new entities WQB and BB is envisaged in terms of QoS negotiation so that QoS-requirements for the whole B3G network domain are provisioned accordingly in the radio access and in the core network as a result of this negotiation. Negotiation of QoS is achieved by a policy-based service negotiation protocol (e.g. COPS-SLS [8]).

Within the proposed QoS framework, the following study items constitute the main building blocks for a consistent end-to-end QoS framework:

- WQB encompassing CRRM functionalities as the key entity for the QoS management in the radio segment.
- BB as the key entity for the DiffServ segment QoS control, further enhanced by QoS routing.

- Mapping optimisation between radio classes (based on UMTS classes) and IP QoS classes (based on DiffServ).

Next subsections provide more details on envisaged solutions for each of the above mentioned aspects. Similarly, the necessary procedures and signalling to implement the proposed QoS framework is developed and described in the last subsection in order to give completeness to the presented solution.

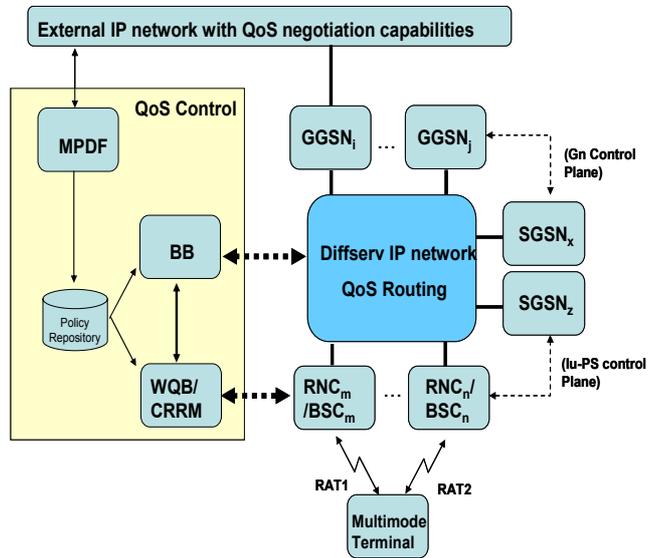


Figure 1. Reference B3G network UMTS-based architecture

III. Common Radio Resource Management

Common Radio Resource Management (CRRM) refers to the set of functions that are devoted to ensure an efficient use of the available radio resources in heterogeneous networks scenarios by means of a proper coordination between the different radio access networks. An inherent CRRM functionality is the RAT selection (i.e. the functionality devoted to decide to which RAT a given service request should be allocated). After the initial RAT selection decision, taken at session initiation, vertical (inter-system) handover is the procedure that allows switching from one RAN to another. Vertical handover procedures from one RAN to another may be useful to support a variety of objectives, such as avoiding disconnections due to lack of coverage in the current RAT, blocking due to overload in the current RAN, possible improvement of QoS by changing the RAT, support of user's and operator's preferences in terms of RANs usage or load balancing among RATs. Thus, the vertical handover procedure enables another dimension into the CRRM problem and provides an additional degree of freedom in rearranging traffic.

In this context, some basic RAT selection approaches are:

- Service-based RAT selection. A service-based RAT selection policy is based on a direct mapping between services and RATs. This includes, as illustrative examples, (1) VG (voice GERAN) policy: This policy primarily allocates voice users into GERAN and interactive services into UTRAN or (2) VU (voice UTRAN) policy: This policy primarily allocates voice users into UTRAN and interactive

services into GERAN. If no capacity is available in the primary RAT, the other RAT is selected instead (thus leading to a combined policy denoted as VG_VU or VU_VG depending on what is the first choice). If no capacity is available in the alternative RAT, the service request gets blocked (at service set-up) or dropped (during service life-time).

b) Load-balancing based RAT selection. Load balancing (LB) is a possible guiding principle for resource allocation in which the RAT selection policy will distribute the load among all resources as evenly as possible.

Figure 2 shows the voice call dropping probabilities (in %) for both VG_VU and LB algorithms, under a UTRAN/GERAN scenario described in [9]. Up to 600 voice users, dropping values are kept sufficiently low. For 800 voice users however, VG reveals higher dropping values than policy LB. On the other hand, Figure 3 shows the average downlink packet delay exhibited by interactive traffic. Results show that VG policies reveal lower delays than with the appliance of policy LB, especially when increasing voice load. Results indicate that the flexibility provided by VHO is capable enough to redirect users to the most appropriate RAT depending on the current network situation, e.g. load, BLER, etc. Therefore, no significant improvement is introduced by trying to balance the load in the system.

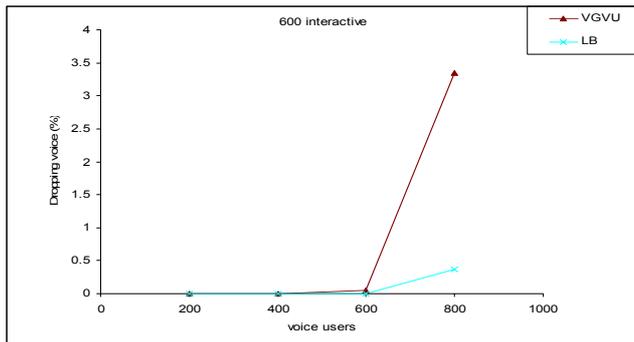


Figure 2. Dropping Probability (%) for voice users considering policy VG_VU and LB with VHO.

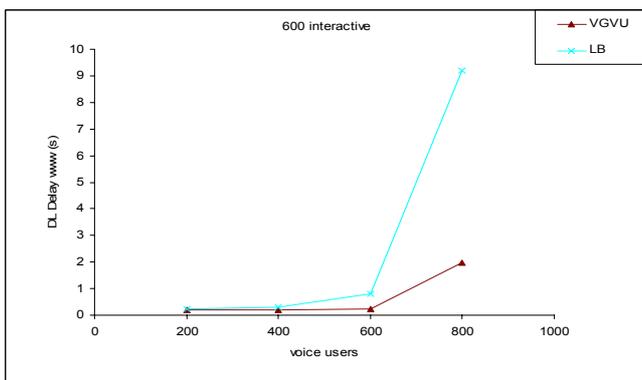


Figure 3. DL Average Packet delay (s) for interactive users considering policies VG_VU and LB with VHO.

IV. Mapping

The notion of end-to-end QoS presupposes the existence of several QoS domains and interactions between them. In the

case of a B3G access network, this is materialized by the interactions between the different RANs and the IP CN domain. These QoS interactions can be categorized into the coordination, on one side between the admission control mechanisms, and on the other side between the QoS data planes, which are fundamentally different in the RAN and the IP CN domain. In the IP CN, DiffServ QoS mechanisms involve traffic aggregates and in the RAN QoS mechanisms consist of per-flow management. This latter coordination involves the mapping of QoS classes in the RAN, aligned on the UMTS RAB with IP CN QoS classes, which are based on DiffServ classes as indicated in [10]. This section examines the end-to-end QoS performance of the static QoS class mapping in some scenarios found relevant in EVEREST because of the important role that a proper mapping design could have in terms of resource optimisation. In particular, scenarios where CRRM deployment could have an important impact on the characteristics and volume of the traffic supported in the access links from Radio Network Controllers to backhaul or backbone network elements are envisaged. Two sets of simulations have been carried out. A first set of simulations examines the QoS class mapping issue at a single AR without considering the effect of the CRRM algorithm on the traffic distribution on the IP CN. Then, another set of simulations, examines the QoS class mapping considering the effect of CRRM algorithm on the traffic distribution in the IP CN.

The considered scenarios are concerned with optimisation of the capacity available in access links from radio network controllers. The deployment of CRRM strategies in the radio access network might significantly alter user distribution between radio network controllers (i.e. RNCs and BSCs) so that capacity provisioned in access links can not suffice for specific traffic patterns. On the other side, the progressive growth of 3G traffic, along with the introduction of high speed data channels in radio access networks, might derive in higher bandwidth requirements in some access links. In those situations, CRRM strategies may be used to distribute the traffic in a way that capacity usage of access links is optimised and so unnecessary capacity upgrades in the transport network are avoided.

Figure 4 illustrates the envisaged situation. As shown, some users can access the B3G network through different radio attachment points (e.g. either available candidate cells are of the same technology but connected to different RNCs or just consider different radio access technologies). Under that situation, the “best” cell should be determined depending on radio aspects as well as IP transport.

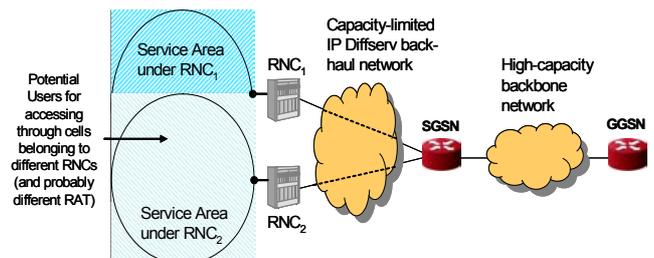


Figure 4. Dependency of traffic patterns in access links on CRRM strategies

As an illustration of this situation, Figure 5 and Figure 6 plot the effects of the two CRRM algorithms “load balancing” (LB) and “service-based policy” (VG_VU) as explained in section IV, on the performance of DiffServ QoS classes. The scenario considers a situation with a high load in the radio interface approximately equally split among a voice and an interactive service. Specifically, the two figures plot the input traffic pattern in the BSC node for the two algorithms. Similar effects not shown here for the sake of brevity would be observed in the RNC. Obviously, as it can be seen in the figures, the LB CRRM algorithm has a load balancing effect on the aggregated traffic flows (voice and interactive) between the RNC and BSC and therefore the BSC serves the two types of service classes. On the contrary, with the service-based CRRM strategy, although the average aggregated throughput is similar than with LB, the traffic classes split is very different. Clearly, this has also an impact in terms of performance as it is shown for example in Table 1 from the point of view of delay.

Table 1. Performance metrics with LB and VG_VU

	VG_VU	LB
Average delay interactive BSC	13.73 ms	17.9 ms
Jitter interactive BSC	0.28 ms	0.65 ms
Average delay voice BSC	17.28 ms	18.45 ms
Jitter voice BSC	0.15 ms	0.33 ms

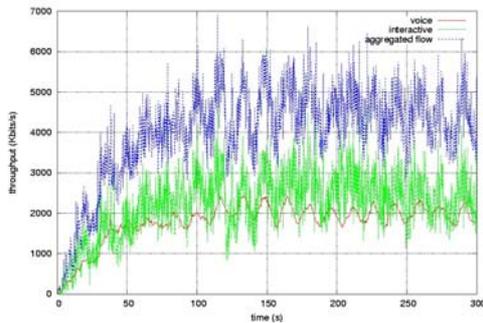


Figure 5. Throughput in the BSC with the LB CRRM approach

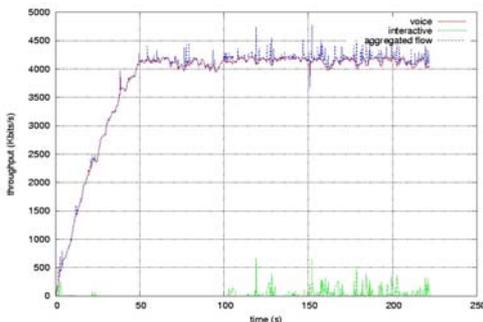


Figure 6. Throughput in the BSC with the VG_VU CRRM approach

V. Bandwidth Broker and QoS Routing

The implications of using the BB concept in an IP mobile access network have been addressed in the EVEREST project. Basically, the BB is an agent that has the information of all the resources in a specific diffserv

domain and that can be consulted in admission control decisions. Decisions taken in the BB can be simply based on purely static bookkeeping with parameter-based admission control but also more elaborated approaches relying on active measurements can be used [11].

It has been shown in EVEREST that the use of BB in an IP mobile access network implies the analysis of the interactions between BB, QoS routing and IP micromobility overlay [12]. In particular, the mechanisms involved in the initial access router selection and IP handover are dependent on the usage of either host-based or tunnel-based micromobility protocol and either source or hop-by-hop QoS routing protocol.

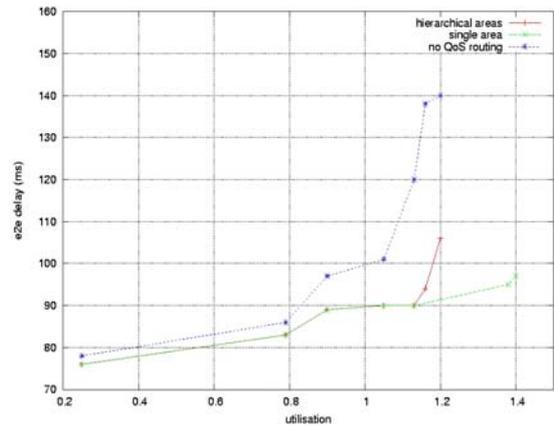


Figure 7. VoIP mean e2e delay versus mean utilisation

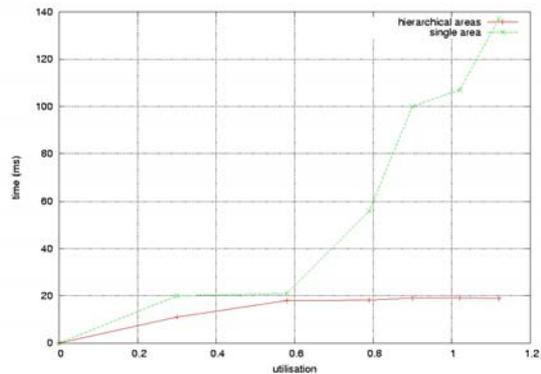


Figure 8. Convergence time for a single QoS routing area and hierarchical areas

As the BB is a centralised approach, considerations about scalability are of prime importance. In that sense, solutions based on hierarchical BBs have been analysed. It has been shown that there is a trade-off between scalability and QoS performance. In particular, it can be stated that a hierarchy of BBs with hierarchical QoS routing areas presents a lower signalling overhead compared to a single BB with a single QoS routing area. Nonetheless a higher network utilisation can be achieved for a single BB with a single QoS routing area. As an example, Figure 7 shows the mean delay of voice packets versus link utilisation in case of using hierarchical areas or just a single domain. It can be observed that both the hierarchical and single-area solutions behave quite similar but the hierarchical approach begins to

offer worse results than a single-area BB for high link utilisation margins. On the other hand, convergence time for the case of using QoS routing in hierarchical areas is considerably lower, as shown in Figure 8, because as the number of routers is smaller, the link state advertisements go through a smaller number of hops. Furthermore it can be noticed that the convergence time in the case of QoS routing is a function of the traffic and therefore also the network utilisation.

VI. Functional Allocation and Procedures

As from the proposed functional architecture, its practical feasibility has been assessed by means of the definition of two key procedures: connection establishment and vertical handover. Both procedures allows for a RAT selection decision. In a heterogeneous radio access network scenario with multimode terminals, a RAT selection mechanism to decide the most appropriate cell or point of attachment to the network becomes mandatory. In a general sense, it can be stated that the RAT selection problem should be managed by the mobile user when available candidate networks are not coordinated (i.e. a user must decide which is the “best” network by himself, and of course, in this decision operator and network constraints are implicitly considered –cost, subscription profile and so on--). However, when some degree of coordination exists in the available candidate networks (e.g. CRRM), the user may want to delegate “part of the decision” to the coordinated network where a better decision is expected to be done in terms of both a better user satisfaction and effective usage of network resources. The meaning here of “part of the decision” is that although RAT selection decision would be taken in the network, user preferences would be considered in the decision process (i.e. from its user profile). Thus, in the composite network environment described above a RAT selection function controlled by the network although based on user preferences is envisaged. As an example of carried work in this aspect, Figure 9 illustrates how a VHO mechanism would be decided and executed in the considered B3G network scenario (for the sake of simplicity only two radio network controllers RNC1 and RNC2 belonging to any of the considered RATs are illustrated in the figure). The basis of the vertical handover mechanism shown here is the Combined Hard Handover and SRNS Relocation procedure provided in 3GPP 23.060. In this case, CRRM entity can be thought to be collocated within the WQB functions. Thus, in this case, initial vertical handover decisions would be triggered at local RRM entities according to installed policies. According to Figure 9, the source RRM entity decides that a VHO may be needed and a list of candidate cells is passed to the CRRM entity in the WQB. The CRRM entity there might use measurements gathered from other RNCs to prioritise the provided list in terms of radio resource usage. However, the decision of moving to a target cell of a different RAT should consider criteria other than merely radio specific. In particular, service type, network operator and user preferences must be included in the decision process. So, the WQB might directly retrieve operator policies from a global repository (that can be the same used by the MPDF when evaluates session establishment, i.e. when the AF asks the MPDF about QoS authorisation) and specific

information from the user databases (e.g. HSS) so as a decision can be taken locally at the WQB. After that, if the handover decision implies a change in the point of attachment to the core network (i.e. a RNC change according to the proposed architecture), the BB should be checked to guarantee that QoS would be maintained at the new path. Concerning the admission control at the IP level, the BB centralizes all the information regarding the current network utilization. As it has been noticed in section 2, QoS routing is used in the IP domain for the determination of the new IP path after handover. Thus, the QoS routing determines the new path along which the required service is available and this information is transmitted to the BB. Further details of both procedures can be found in [13][14].

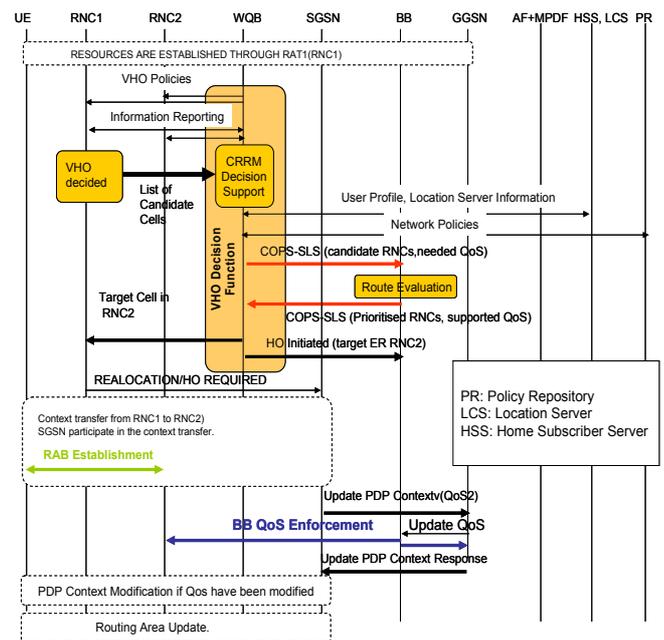


Figure 9. Message chart for a Vertical Handover procedure

VII. Towards End-to-End QoS management in B3G networks: AROMA’s vision

AROMA project is expected to leverage EVEREST’s achievements so that a progressive transition towards an all-IP radio access network is pursued. In this sense, as illustrated in Figure 10, a first step scenario considered in AROMA accounts for the introduction of IP transport in the radio access network. Based on envisaged benefits of IP-based RAN networks (e.g. lower costs, flexibility of merging wired and wireless networks, and network scalability and reliability), it is also expected that wireless operators may gain significant OPEX reductions thanks to the support of alternative transmission solutions (e.g. Ethernet services). Therefore, it is necessary to develop transport techniques that can meet QoS requirements in the UTRAN as well as exploit the potential interactions between radio resource management and QoS mechanisms deployed in the transport layer.

With this first scenario, AROMA is mainly concerned with the optimisation of end-to-end QoS mechanisms encompassing common radio resource management functions in the radio access part and IP QoS mechanisms

in the core and radio access transport networks, while maintaining interoperability with the legacy network. Then, in a second stage, AROMA will consider new architecture proposals aligned to 3GPP Long Term Evolution activities where network architectures are moving from previous hierarchical UMTS network, influenced by GSM and ISDN history, towards decentralized and distributed architectures based on IP principles. This approach leads to re-think the network as a whole in order to provide the most efficient solution. In that sense, AROMA is expected to provide some valid approaches to handle these new architectures where overlapping between radio and transport resource management becomes more evident and proper coordination is a challenge.

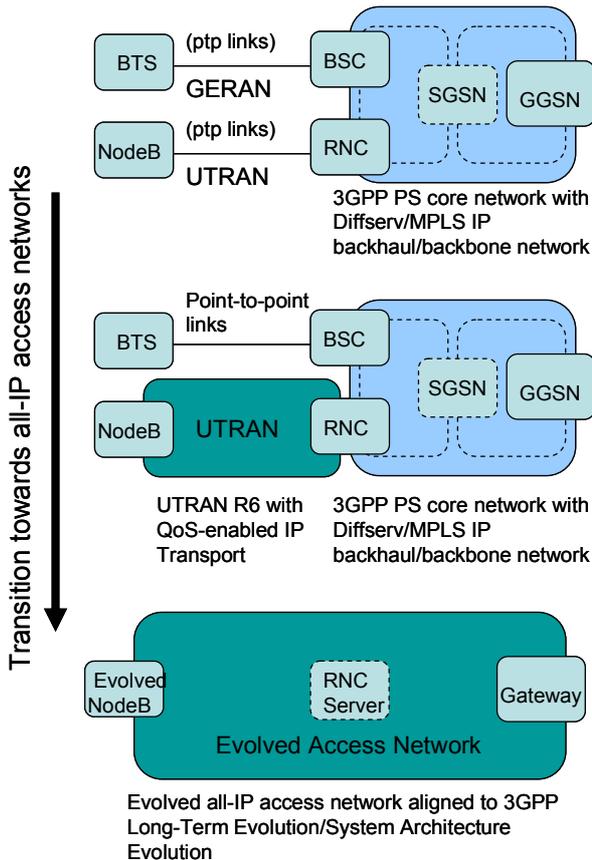


Figure 10. Transition to an all-IP access network

VIII. Conclusions

One important research activity within the IST-AROMA is devoted to the study of innovative end-to-end QoS strategies considering both radio and transport network aspects in the context of an all-IP heterogeneous access networks. In this paper, the main building blocks of the envisaged QoS framework have been identified along with some architectural aspects of the all-IP network under consideration. Results achieved in the IST-EVEREST project have been provided in order to support the proposed QoS framework.

ACKNOWLEDGEMENTS

This work is part of the EVEREST and AROMA projects, partially funded by the European Commission under the IST framework. The Authors would like to acknowledge

the contributions of their colleagues from the EVEREST and AROMA projects.

REFERENCES

- [1]. X. Fu, H. Karl, and C. Kappler, "QoS-Conditionalized Handoff for Mobile IPv6", *Networking2002*, LNCS, vol. 2345, pp. 721-730, 2002.
- [2]. J. Kakishima et al., "'Plug and Access" for heterogeneous radio access environment", *WWRF10 New York*, Oct. 2003. [Online]. Available: <http://www.wireless-world-research.org>
- [3]. B. Moon and H. Aghvami, "Seamless Switching of RSVP Branch Path for Soft Handoff in All-IP Wireless Networks", *IEICE TRANS. COMMUN.*, vol. E86-B, no. 6, pp. 2051-2055, June 2003.
- [4]. N. Nafisi et al., "Extending QoS policy-based mechanisms to B3G mobile access networks", *IST Mobile Communications Summit*, June 2004.
- [5]. 3GPP TS 23.207, "End-to-end Quality of Service (QoS) concept and architecture (Release 6)".
- [6]. R. Neilson et al., "A discussion of bandwidth broker requirements for Internet2 Qbone", available online on http://www.merit.edu/i2qbone-bb/doc/BB_Requirements_v3.doc.
- [7]. A. Campbell et al., "Comparison of IP micromobility protocols", *IEEE Wireless Communications*, 2002.
- [8]. T. Nguyen et al., "COPS-SLS usage for dynamic policy-based QoS management over heterogeneous IP networks", *IEEE Network Magazine*, May 2003.
- [9]. P.Karlsson et al. "Target Scenarios specification: vision at project stage 2", Deliverable D13 of EVEREST project, available at <http://www.everest-ist.upc.es>
- [10]. 3GPP TS25.414 "UTRAN Iu interface data transport and transport signalling".
- [11]. J. Lakkakorpi, "Simple measurement-based admission control for DiffServ access networks", in *Proc. ITCOM 2002*, July 2002, Boston, SPIE.
- [12]. N. Nafisi et al. "Final report on QoS management in the Core Network and QoS Mapping", Deliverable D16 of EVEREST project, available at <http://www.everest-ist.upc.es>
- [13]. R. Ferrús et al. "A Feasible Approach for QoS Management in Coordinated Heterogeneous Radio Access Networks", *24th IEEE International Performance Computing and Communications Conference (IPCCC 2005)* – April 2005 - Phoenix, Arizona
- [14]. R. Ferrús et al. "Vertical Handover Support in Coordinated Heterogeneous Radio Access Networks", *14th IST Mobile & Wireless communications Summit 2005*, Dresden, Germany.