

User perceived Quality Evaluation in a B3G Network Testbed

Xavier Revés, Nima Nafisi, Ramon Ferrús, Antoni Gelonch

Abstract—This paper provides some experimental results on end-user perceived quality obtained in a beyond 3G (B3G) network testbed developed within the EVEREST project. The EVEREST testbed has been designed to demonstrate some of the main concepts addressed within the project, concerning both: Common Radio Resource Management strategies and end-to-end QoS architectures and mechanisms for B3G systems based on the UMTS architecture. The complexity of the interaction between B3G systems and the user applications, while dealing with the QoS concept, pushes to develop this kind of emulation platforms, where algorithms and applications can/must be tested in realistic conditions, not achievable by means of off-line simulations. Addressed results show how some CRRM strategies for RAT selection can impact on voice quality perception and the quality degradation of a video streaming session caused by the disruption time occurring when a vertical handover is performed..

Index Terms—beyond 3G, DiffServ, end-to-end QoS, heterogeneous access network, policy-based service negotiation

I. INTRODUCTION

THE main objective of the EVEREST project has been to investigate and propose mechanisms and algorithms that can handle the expected traffic growth and the more demanding Quality-of-Service (QoS) services in a heterogeneous network structure which comprises 2G and 3G cellular systems as well as wireless local area networks (WLANs). The reference architecture is aligned with the work carried out in 3GPP and IEEE 802.11. Furthermore, the selected scenarios and the corresponding evaluation procedures are compliant with 3GPP specifications in order to facilitate the impact of the project result in 3GPP.

Due to the complexity of such heterogeneous systems, an appropriate definition of the envisaged scenarios is crucial in order to determine the performance of algorithms and strategies, especially when a manifold of users, services and radio access technologies are involved. The considered scenarios are mainly based on the requirements and visions of the four Mobile Operators that participate in the project,

This work is part of the EVEREST project, partially funded by the European Commission under the IST framework and by the Spanish Research Council under TIC2003-08609 grant.

Xavier Revés, Ramon Ferrús and Antoni Gelonch are with the Signal Theory and Communications Department of the Polytechnic University of Catalunya, Barcelona (e-mail: [xavier.reves, ferrus, antoni]@tsc.upc.edu).

Nima Nafisi is with Centre for Telecommunications Research, King's College, London (e-mail: nima.nafisi@kcl.ac.uk).

which are interested in analyse the impact that resource management algorithms have on the system performance. Then, such scenarios should consider different radio networks capabilities, traffic load conditions, propagation characteristics, user mobility patterns, service configurations for the all mobile users, as well as different deployments, and network configurations where several radio access networks (RANs) can give access to offered services through a common core network (CN).

In such context, EVEREST testbed is devoted to demonstrate the benefits of the developed Common Radio Resource Management (CRRM) algorithms and proposed QoS management techniques. Basically this demonstration framework reproduces, in real time, the conditions that the wireless heterogeneous network behaviour, including the effect of the other users, produces over the user under test (UUT), when making use of real IP-based applications (i.e. videoconference). Then, the EVEREST testbed consist of a GERAN/UMTS/WLAN stand-alone real time emulator platform, including all the relevant QoS entities in both the radio access part and the CN, to show and analyse the end-to-end QoS performance.

Summarising, the EVEREST testbed aims to provide a number of features that are not easily achievable by means of conceptual studies or system level simulations. Among such features, it should be emphasised the possibility to test the end-to-end QoS performance and to assess, in real time, the impact that RRM/CRRM/Bandwidth Broker (BB) algorithms, [8][9], have on the user's perceived QoS.

II. TESTBED DESCRIPTION

As the EVEREST testbed has been designed to demonstrate most of the concepts addressed within the project [2], its architecture has been defined taking into account the EVEREST project reference architecture, [1][3][4][5], and the required procedures to validate such concepts under the defined scenarios.

A. Testbed Procedures

1) *Initial RAT (Radio Access Technology) Selection / RAT Switching.* To demonstrate the decision of selecting a specific RAT, including changing from one RAT to another, when several radio access networks are available

2) *Connection establishment with E2E QoS negotiation / re-negotiation.* Complete procedures to establish an end-to-end connection with QoS assurances, and its re-negotiation when

initial conditions cannot be fulfilled, highlighting the decision process and its dependence with the different scenario configurations.

3) *CN Mobility Management and QoS interactions*. This aims to demonstrate the interactions between the IP mobility management, [6], and the IP QoS entity (i.e. the BB) of the CN. These interactions are needed in order to assure the provision of QoS along the new path after a handover.

4) *Common Radio Resource Management operation*. To demonstrate the decision process of the CRRM strategies considered in a given scenario.

5) *Impact of CRRM and QoS Management on Applications* which has been designed to demonstrate the effects of the developed RRM and CRRM algorithms and QoS management strategies over real applications.

B. Testbed Architecture

A detailed view of the EVEREST testbed can be found in [10][11]. Its main components, as shown in Fig. 1, are the following:

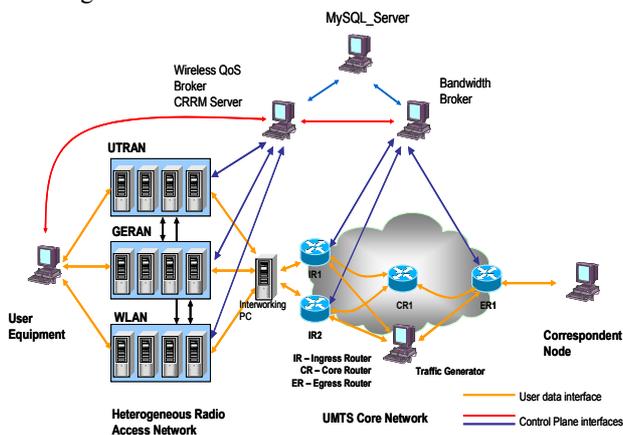


Fig. 1. EVEREST Testbed Architecture

- User Equipment Emulation which hold applications and a QoS client to manage connections through the heterogeneous access network.
- RAN Emulators. A set of emulation platforms to cope with the main characteristics of the UTRAN, GERAN and WLAN technologies has been developed.
- Wireless QoS Broker, that handles QoS management in the heterogeneous RAN as well as CRRM functions.
- Switching Node, used to establish different configurations of coupling between RANs and the correspondent routers in the CN.
- Core Network Routers. The Core Network is based on real enhanced IP routers implemented over PCs with Linux. Diffserv mechanisms have been included to provide QoS.
- Bandwidth Broker which acts as the Core Network QoS management entity [7].
- Traffic Generator. Node in charge of generating controlled traffic to properly load the core network.
- Data Base based in MySQLServer used as policy repository.

- Correspondent Node which holds applications.

III. END-USER PERCEPTION QUALITY EVALUATION

Nowadays, significant research is being performed in the area of end-user perceived quality assessment and two main directions are being explored: objective and subjective testing of perceived QoS.

A. Subjective Evaluation of QoS

The methodology to capture the subjective perception of quality when using multimedia applications should be based in some kind of MOS (Mean Opinion Score). Taking into account that the most representative multimedia applications, streaming and videoconference, are a combination of video and audio it can be assumed that its evaluation methodology can be quite similar. The MOS protocols for video and audio QoS evaluation are specified in the ITU-T (International Telecommunication Union) where the typical schemes uses subjective tests (opinionated scores) that are mathematically averaged to obtain a quantitative indicator of the system performance. To determine MOS, a number of listeners/viewers rate the quality of video and sound transmitted through a communications system. The MOS is the arithmetic mean of all the individual scores, and can range from 1 (worst) to 5 (best). In document P.910 [12], the main recommendations regarding MOS protocols/methods are described.

B. Objective Evaluation of QoS

Although the previously described methodologies to measure the perceived quality are the traditional ones the complexity and cost of such test makes them not feasible for an automatic evaluation procedure. Some new procedures have been developed based in the use of some algorithms that are able to provide the estimation of the perceived quality from some objective measurements. Each type of information, video, voice, etc. has its own algorithm and methodology. In case of the quality of the perceived voice/sound the available perception model standardised by the ITU is known as PESQ (Perceptual Evaluation of Speech Quality) [13] is an enhanced perceptual quality measurement for voice quality in telecommunications. In case of video, the most extended method is based in the computation of the PSNR (Peak Signal to Noise Ratio) for each image in the video stream and computing the distortion in a pixel basis. The PSNR method compares the maximum possible energy of the signal with the noise energy. Nevertheless, the ITU-T recommendation J.144 rev. 1, also known as VQM (Video Quality Metric) [14], defines objective computational models that have been shown to be superior to PSNR as automatic measurement tools for assessing the quality of video.

C. Audio: PESQ tools

Determining the subjective speech quality of a transmission system has always been an expensive and laborious process. The PESQ tool, incorporated to the EVEREST testbed, provides a rapid and repeatable result in a few moments

forecasting automatically the quality score obtained from a subjective test. The basis of such test is the modelling of the human audition perception capturing the quality of sound/voice by comparing the original signal $X(t)$ with the degraded signal $Y(t)$ obtained as result of transmitting the signal $X(t)$ over the communication system under test. The output of the PESQ is the forecast of the averaged perceived quality of a set of participants in a subjective test of the signal $Y(t)$.

D. VQM (Video Quality Metric) Tool

The VQM Software incorporated to the EVEREST testbed provides a means of objectively and repeatedly evaluating video quality. The software compares an original video and a processed (sent through system under test) video and reports a Video Quality Metric that correlates to the perception of a normal end user. The VQM estimation methods are based on processing 8-bit digital component video as defined by Recommendation ITU-R BT.601. The encoder can utilize various compression methods (e.g. MPEG, H263, etc). The models proposed in the J.144 Recommendation may be used to evaluate a codec (encoder/decoder combination) or a concatenation of various compression methods and memory storage devices. Each of the VQM models is a particular algorithm used to compute VQM that has been specifically optimised to achieve maximum objective to subjective correlation based upon certain optimisation criteria, including the range of quality over which the model applies and the speed of computation. They are described in detail in the NTIA Report 02-392 [16].

IV. RESULTS

Results provided below are organised in two trials. The first trial is addressed to show how some CRRM strategies for RAT selection can impact on voice quality perception. Then, the second one is aimed at assessing the estimation of the quality degradation of a video streaming session caused by the disruption time due to a vertical handover execution.

A. Trial 1. CRRM policies and voice quality assessment

In this trial, the testbed is configured to emulate a service area where both UTRAN and GERAN are available. Then, two different CRRM policies concerning the allocation of voice users to UTRAN or GERAN are analysed. In the first policy, referred to as “random” selection policy, a voice session can be handled indistinctly through UTRAN or GERAN when both networks have enough resources to do it. On the other hand, in the second policy referred to as “indoor GERAN”, GERAN selection is prioritised over UTRAN for those connections with higher path losses (e.g. indoor traffic). The rationale behind the “indoor GERAN” policy is based on the fact that connections with higher path losses lead to a higher performance degradation in UTRAN than in GERAN [17]. Therefore, next results demonstrate how the performance degradation produced by indoor users in UTRAN can be alleviated by diverting such traffic to GERAN.

The admission control procedure implemented in the testbed allows distributing the users among the available RATs according to network usage, user preferences and operator preferences. A functional organisation of this admission control is depicted in Fig. 2.

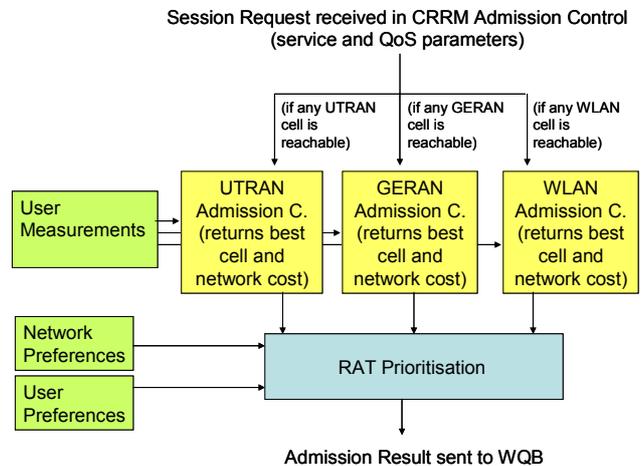


Fig. 2. RAT Selection algorithm implemented in the testbed

When a new user session request arrives, local admission entities in every RAN are checked in order to identify where the new user could potentially be connected. Local admission control results in a prioritisation of the candidate cells of a given RAT and also in an indication of the network cost, in terms of resource consumption, associated to such decision. Once admission information from local admission control is passed, a prioritisation is done among those cells where the user sessions could be established. So, a priority value is obtained for each available RAT by using the following expression:

$$\begin{aligned} \text{Priority}(\text{RAT}, \text{service}) = & \\ & W_{\text{NP}} * \text{NP}(\text{RAT}, \text{service}) + W_{\text{UP}} * \text{UP}(\text{RAT}, \text{service}) \\ & + W_{\text{RC}} * (100 - \text{RC}(\text{RAT}, \text{service})) \end{aligned}$$

where:

- $\text{UP}(\text{RAT}, \text{service})$: User Preferences policies to use a given RAT for a given service. Its value ranges from 0 to 100%.
- $\text{NP}(\text{RAT}, \text{service})$: Network Preferences policies to use a given RAT for a given service. Its value ranges from 0 to 100%.
- $\text{RC}(\text{RAT}, \text{service})$: Resource Cost when using a given RAT for a given service. Its value ranges from 0 to 100%.

and W_{NP} , W_{UP} and W_{RC} are the weights (also from 0% to 100%) used to account for network preferences, user preferences and resource consumption respectively.

In this demonstration only voice traffic is considered. In particular, 1200 voice users are uniformly distributed in a 8Km x 4Km scenario served by 13 UTRAN/GERAN co-sites. UTRAN base stations operate a single FDD carrier while GERAN BTS use 3 carriers. Then, 50% of the users in the scenario are considered to be outdoor users moving at 3km/h

while the rest are indoor users. Additional propagation loss for indoor users is assumed to be 15dB. Under such conditions, Table I provides results obtained for both CRRM policies in terms of admission statistics. As observed in the table, under both policies, global admission rejections are kept very low for the considered scenario settings. So, from a service availability perspective, both policies behave quite similar. Then, regarding RAT selection, when random selection is enforced, UTRAN is chosen 61% of the times while this percentage drops to 58% when specific policy for indoor users is applied. The reason why UTRAN is in both cases the most used RAN is because of its higher capacity. This can be checked by observing the local admission rejections in both networks: while local admission control in UTRAN is ready to accept all the sessions being requested, local admission control in GERAN blocks a high percentage of them (ex. 28.7% with the “random selection” policy and 19.1% with the “indoor users to GERAN” policy). Also it is interesting to notice that UTRAN coverage is 98.5% in the analysed scenario while GERAN has a 100% coverage.

TABLE I
ADMISSION RESULTS

Radio-based policy	RAT Selection (UTRAN/GERAN)	Global Admission Rejections (%)	Coverage Availability (UTRAN/GERAN)	Local admission rejections (%) (UTRAN/GERAN)
Random selection	61% / 39%	0,42%	98,5% / 100%	0% / 28.7%
Indoor GERAN	58% / 42%	0,26%	98,5% / 100%	0% / 19.1%

But although voice sessions are accepted in the system with a low rejection probability in the two cases, resource consumption and quality of service parameters is quite different for those sessions handled by UTRAN. As observed in Fig. 3, “random selection” policy results in a higher packet loss ratio due to a higher interference level as well as mobile terminal power limitations.

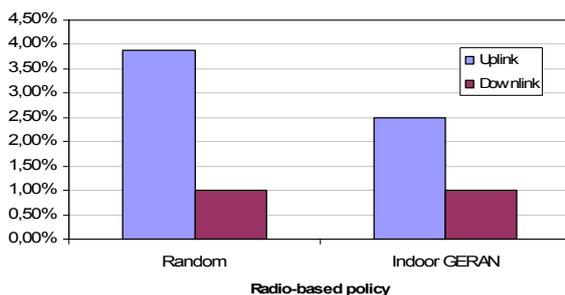


Fig. 3. Voice packet loss ratio in UTRAN

Upon such conditions, user perceived quality is evaluated by establishing a real voice call through the testbed. In particular, the rat audio conferencing software has been used to transmit several times the same recorded audio file. The results of the evaluation of the MOS quality by means of the PESQ software tool are reflected in the Table II.

The figures in Table II have been obtained with an audio

source one minute long transmitted in the downlink with a channel requested bandwidth configured to 128kbps and an audio transmission rate lower than 100kbps. Several different trials per case have been performed and evaluated to obtain a mean quality measurement. To minimise the effect of the recording error at reception, the quality has also been normalised with that obtained when the channel does not introduce any effect.

TABLE II
QUALITATIVE MEASUREMENTS OF AUDIO

Lose ratio (%)	Quality: 1 (Worse) – 5 (Best)	
	Mean	Relative quality (%)
0	3.3	100
1	3.1	94
2	3.0	91
3	2.9	88
4	2.8	85
5	2.7	82

It is worthy to notice that the effect of a higher packet error ratio has a lower impact on the quality of the audio than that shown in the case of video streaming. However, an increment of one percentage point in the packet error ratio translates in a three points reduction of the audio quality.

B. Trial 2. Disruption Time

The goal of this demonstration is to evaluate the impact of the change of IP attachment point on the application used by the UUT. At the moment of IP handover, signalling between mobility management entities and also QoS entities is involved; therefore it can create a short period of interruption for the application. This effect is evaluated here by means of objective QoS perception assessment for a video-streaming application.

To isolate the effect of the disruption time on the application, the scenario is configured to produce a disruption time periodically because of having performed a vertical handover. Both RATs involved in the vertical handover are not highly loaded to avoid any kind of restriction in the traffic that flows through them, so the main observed effect will be produced by the disruption time. Two parameters are modified to obtain the results shown in the following section. First of all, the time interval between two consecutive announcements (routing advertisement) of the routing software located at every access router to make the mobile node to notice the new router availability. The second parameter to modify is the policy applied in the traffic switch entity when a vertical handover is made, which can be either removing the accumulated packets (discard policy) in the buffers or sending them (forward policy) using the new RAN.

To measure the effect on the quality perceived by a user with an application that suffers the disruption of the network connectivity, the previous tests have been repeated using a tool that can provide a qualitative measurement of the quality of a video. A video of 1 minute length is transmitted through a streaming service with two disruptions during this minute.

Different trials per value of the disruption time are performed. For every one, the received video is captured and compared against the video received when no disruptions happen. The video quality measurement tool (VQM) retrieves the qualitative measurements for every case, which are listed in the Table III for three independent tests.

The conditions of the tests are the following:

- Downlink requested bandwidth: 150kbps
- Uplink requested bandwidth: 32kbps
- Video coding rate: ≤ 128 kbps
- Streaming protocol: RTSP
- Streaming player: QuickTime
- Streaming server: DarwinStreamingServer

With a higher data rate in the downlink it is possible to ensure that no packets will be discarded because of buffer limitations (per-RAN configurable). The 150kbps value is adjusted to allow transporting the video and audio data over RTSP without lost of packets because of the buffer capacity for the one minute length test. Because there is not a large bandwidth excess, when packets are discarded in the buffers a sufficient amount of data is lost to produce an unrecoverable lost of quality in the video. Increasing the bandwidth reduces the amount of packets lost during the vertical handover (VHO) because fewer packets are stored in the buffers. Then, a large bandwidth excess reduces the effects of discarding packets.

TABLE III
QUALITATIVE MEASUREMENTS OF VIDEO STREAMING

Disruption Time (seconds)	Buffer Policy	Quality: 1 (Worse) – 5 (Best)			
		Test A	Test B	Test C	Mean
1	Discard	<1.5			
	Forward	3.90	3.46	4.7	4.02
4	Discard	<1.5			
	Forward	3.59	3.1	4.17	3.64
7	Discard	<1.5			
	Forward	2.96	3.53	4.21	3.56
10	Discard	<1.5			
	Forward	3.48	2.92	3.40	3.26

In any presented case, and taking into account the previous conditions, discarding the packets accumulated in the buffers results in a dramatic lost of quality. The video stalls, the images combine each other and the colours are corrupted after the first disruption. Before this initial disruption happens, the quality is acceptable, giving a global quality result higher than 1 (worse) but, in any case, actually very poor from the viewpoint of human quality perception. In the table, it just appears an indication that the quality is always observed as extremely poor (Mean Opinion Score (MOS) < 1.5).

V. CONCLUSION

In this paper the main features of the testbed implemented in the EVEREST project has been described. Starting from an overview of the architecture model, the main functions and procedures deployed in the testbed has been described, especially those related with the common radio resource

management strategies and end-to-end QoS issues, which are the main concepts that the EVEREST project is focused in. Details of the EVEREST testbed architecture have been provided, including the different modules and interfaces and the methodology followed in their implementation. Special emphasis has been done in the approaches considered for the RAN and CN parts for taking into account all the relevant aspects of the end-to-end QoS concept and the methodology followed to capture the QoS perceived by the end user when using relevant applications.

REFERENCES

- [1] N. Nafisi et al., "Extending QoS Policy-based mechanisms to B3G Mobile Access Networks", *IST Mobile & Wireless Communications Summit 2004*, Lyon, France, June 2004.
- [2] Deliverable D05 EVEREST project "Target Scenarios specification: vision at project stage 1". Available: <http://www.everest-ist.upc.es>.
- [3] G. Fodor et al. "Providing Quality of Service in Always Best Connected Networks", *IEEE Communications Magazine*, July 2003.
- [4] Wei Zhuang, et al. "Policy-Based QoS Management Architecture in an Integrated UMTS and WLAN Environment" *IEEE Communication Magazine*, November 2003.
- [5] Ramon Ferrús et al. "End-to-End QoS Architecture for Multi-Domain and Wireless Heterogeneous Access Networks: the EVEREST approach", Contribution to the World Wireless Research Forum (WWRF) meeting held in Beijing, China, February 2004.
- [6] C. Boukis, N. Georganopoulos, and H. Aghvami, "A hardware implementation of BCMP mobility protocol for IPv6 networks", *IEEE Globecom*, vol. 6, pp. 3083-3087, Dec. 2003.
- [7] S. Jha et al., "Java implementation of policy-based bandwidth management", *International Journal of Network Management*, vol. 13, no. 4, pp. 249-258, July 2003..
- [8] 3GPP TR 25.922 v4.0.0, "Radio resource management strategies".
- [9] 3GPP TR 25.881. "Improvement of RRM across RNS and RNS/BSS".
- [10] Deliverable D12 EVEREST project "Implemented Test-bed: Subsystems and Modules". Available: <http://www.everest-ist.upc.es>.
- [11] Deliverable D18 EVEREST project "Integrated Test-bed". Available: <http://www.everest-ist.upc.es>.
- [12] ITU-T Rec. P. 910, "Subjective video quality assessment methods for multimedia applications" (1999).
- [13] ITU-T 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".
- [14] ITU-T Recommendation J.144 rev. 1: "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference".
- [15] V. Bhaskaran and K. Konstantinides. *Image and Video Compression Standards*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995
- [16] Stephen Wolf, Margaret Pinson. "Video Quality Measurement Techniques". NTIA Report 02-392.
- [17] O. Sallent (editor) et al. "Final Report on the evaluation of RRM/CRRM algorithms", Deliverable D20 of the EVEREST IST-2002-001858 project, October, 2005. Available at <http://www.everest-ist.upc.es/>.