Experimental Test-bed for VoIP/VoWLAN Voice Quality Measurements

ROBERT BLATNIK*, GORAZD KANDUS+, TOMAŽ JAVORNIK+
*Department of Intelligent Systems, +Department of Communication Systems
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana
SLOVENIA
robert.blatnik@ijs.si, http://dis.ijs.si,
gorazd.kandus@ijs.si, tomaz.javornik@ijs.si, http://www-e6.ijs.si/

Abstract: In the following paper, a low cost test-bed for evaluating the VoIP/VoWLAN performance is presented. The test-bed is designed and built applying commercial off-the-shelf equipment and for the most part, the freeware software. Voice quality testing is based on Personal Evaluation of Speech Quality algorithm enabling different test setups and system configurations.

Key-Words: Voice Quality Testing, PESQ, VoIP,VoWLAN, 802.11a, WINDECT

1 Introduction

Two parallel wired infrastructures currently exist within home and enterprise networks, namely, the old twisted pair cable infrastructure enabling low data rate communications with guaranteed quality of service (QoS) and the wired infrastructure for computer networks usually providing high data rate but the best-effort data services. Both of them are slowly converging into one network based on internet protocol (IP) and thus providing different services with various QoS. The wireless access to PSTN (public switched telephone network) based on DECT, CT2 and many non-standardized solutions, introduced a micro mobility at our homes and companies. After the success of wireless phones in the past two decades, customers expect the same functionality and QoS also in all IP driven networks and as a result, the Voice over Wireless LAN (VoWLAN) is the next milestone to the converged home and enterprise networks [1].

The voice communications play an important role in today's people interaction, hence one of the main objectives for technology developers and network operators is to design the wireless communication systems which provide the expected QoS to their customers. The WINDECT project, which is funded by the Sixth Framework Programme, is aimed at merging the wireless LAN and DECT telephony networks by integrating the professional quality telephony into WLANS [1]. In order to test the voice quality (VQ) within a project, the experimental test-bed for VoIP/VoWLAN voice quality measurement are designed and built.

The voice quality can be measured from user perspective "perceived – subjective" or "objectively measured" with the parameters like delay, jitter and packet loss. The traditional method for subject measurement of VQ is to calculate a mean opinion score (MOS) defined in ITU-T Recommendation P.800.1 [3], while the objective measurements are specified in ITU-T Recommendations P.862 defining "Perceptual evaluation speech quality (PESQ)" [5] and in ITU-T G.107 Recommendation [6].

There exist many professional tools for metrics-based objective measurements. They can be classified in the following two classes: hardware/software test instruments and software-only test packages. However, due to limited financial resource within WINDECT project, we built an experimental VoIP/VoWLAN test-bed, which is based on commercial off-the-shelf (COTS) products and the professional software package OPERA [7], which enables VQ testing.

The article is organized as follows: after the afore stated introduction, a short description on WINDECT – all IP converged network – follows. In the most important part of the article, the test-bed architecture is explained. The test-bed parameters and test scenarios are listed in the next session. After that an example of VQT measurement is shown and results of the VQ measurements are discussed. The paper concludes with remarks offered for further work.

2 WINDECT system

Two parallel wired infrastructures, one with low data rate providing guaranteed QoS mainly used for plain old telephone service and other one broadband offering best effort service mainly applied for computer communications slowly, converge to one all-IP wired infrastructure, where the old telephone service is replaced by the Voice of IP (VoIP) service with the quality compared to the traditional telephone service. However, wireless access indicates a much slower convergence.

A step forward is a WINDECT project, which aims at merging both networks by integrating the professional quality wireless telephony into WLAN, thus trying to guarantee the QoS compared to wireless telephony specified in DECT standards and also unchanged functionality of physical and MAC layer of the WLAN. These can be achieved by merging the upper layers of the DECT protocol stack with current WLAN physical and MAC layer, using a protocol adaptation layer (PAL) [2]. There exist many VoWLAN products on the market, but all of them exhibit the following drawbacks: there is no VoWLAN standard and the products are thus incompatible, the power consumption is still high compared to DECT terminals, the coverage area is small, poor or even no handover, large occupied frequency bandwidth, low number of simultaneous telephone calls. Consequently, QoS is far from expected since most products are operating at crowded 2.4GHz ISM frequency band. To overcome mentioned drawbacks, the WINDECT is addressing the following problems: to design a protocol adaptation layer with the QoS support thus supporting the seemless handover between different access points (AP), to increase the coverage of the WLAN applying emerging technologies in physical layer, such as multiple input/output systems and finally, to propose solutions so as to cope with packed loss, bit errors and delays which may significantly decrease the speech quality.

In a converged enterprise all-IP network, depicted in Fig. 1, which is proposed by WINDECT project [2], data and telephony share the same wired and wireless links. The wired network is based on internet protocol. The access point has to encapsulate the DECT voice data and DECT signaling into IP packets and transfer them to Soft PBX. The DECT protocol stack and session initiation protocol (SIP) or/and H323 protocol are implemented in Soft PBX. Because no standardized protocol for encapsulating the DECT packet exists, a straightforward way is to encapsulate the DECT voice packet into VoIP packets. However, DECT signaling has to be maintained in separate packages. Since the IP network is not synchronous, only the packet synchronization can be achieved, and as a result the quality of speech is reduced.

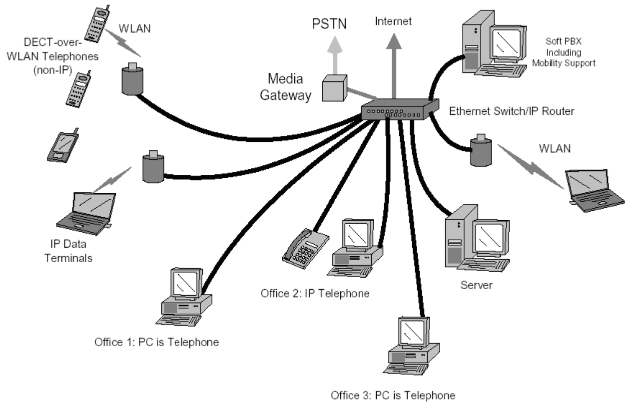

Fig. 1: WINDECT converged all-IP network

In such converged enterprise networks, the quality of voice communication over WLAN is degraded by impairments introduced on air and also by the background traffic competing for the same communication medium (for example, IP data terminal and WLAN Telephone). Therefore, VQT must be performed over a range of background burst and streaming loads as much as possible similar to the real-life traffic demands.

3 Experimental VoIP/VoWLAN test-bed

The metrics based testing has to determine voice-quality performance of VoIP systems in a controlled and repeatable fashion with as much automation support as possible. The measurements are focused on how VQ is determined from the measurements of delay, jitter and packet loss, and how the VQ is affected by voice and data background traffic load, by WLAN signal strength, and roaming. Ideally, VoWLAN VQT should be carried out with so-called "open-air" measurements at the locations of actual VoWLAN systems. This, however, is not feasible in practice, due to a number of technical and economical reasons such as uncontrollable RF interference and high workload for test execution. Repeatable VoWLAN test results can only be achieved in an environment with tightly controlled RF emission, propagation and reflection. Therefore, the VoWLAN voice-quality testing is usually carried out in RF-shielded chambers even though this limits the mobility of WLAN devices under test. Consequently, the testing of mobile terminal handover is limited to emulation of motion, based upon modification of relative RF signal strengths of APs – ideally under software control.

An experimental test-bed, designed and implemented within WINDECT project, is depicted in Fig.2. It consists of several generic building blocks:

- backbone LAN with test manager and burst traffic generator,
- voice signal generator and recorder,
- fixed phone setup,
- mobile phone setup,
- background traffic generator, APs with extra external antennas,
- RF signal attenuators and RF isolated chambers and
- test-bed management part.

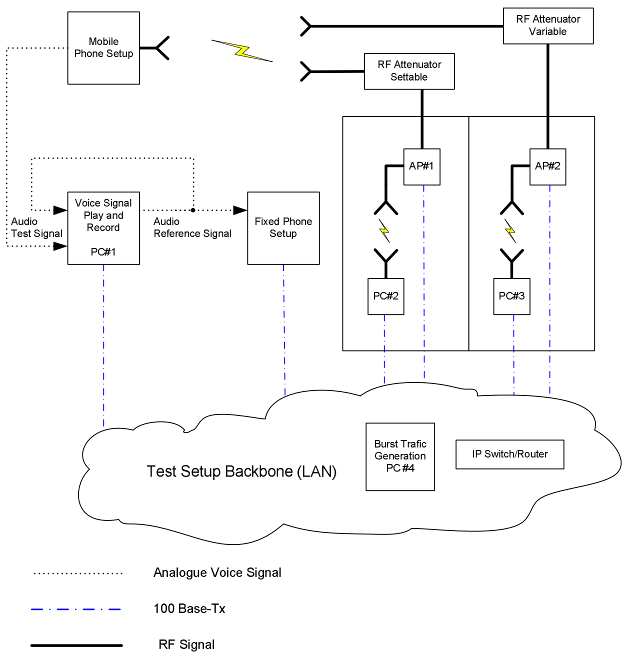

Fig. 2: The VoWLAN VQT Setup

The architecture of the test-bed, shown in Fig. 2 provides four data paths: two voice signal paths (one over currently active AP and one over the AP after handover), and two for the background traffic (one for the burst-type and the other one for the voice-type). The background voice traffic runs between PC#2 and PC#3. Initially, the background traffic was generated by transferring very large files between an FTP server in the test setup backbone and two background-traffic PCs, while later on UDP background traffic was added using traffic generator.

Prior to handover, the VoIP connection from the Fixed Phone Setup traverses the LAN and reaches the Mobile Phone Setup at the other end of the voice connection over the AP#1. Handover transferring the association of the Mobile Phone Setup from AP#1 to AP#2 is triggered by decreasing the RF received power at AP#1 below its sensitivity and/or by sufficiently increasing RF level at AP#2.

The voice signal play/record setup consists of a PC (PC#1) running software for playback of pre-recorded digital voice signal files and feeding this signal into the fixed phone setup. Additionally, the same signal is fed also into one of the two audio inputs of the PC#1 as the reference voice channel of the test file recorded for off-line VQ processing. The fixed voice phone setup converts the reference voice signal into UDP/IP packet stream which is delivered over test-bed backbone LAN to the two APs. From there, this packet stream is transmitted over WLAN to the mobile VoIP phone segment applying selected external antenna on APs. The AP is selected according to the setting of the RF signal attenuators (RFA#1 and RFA#2 in Fig. 2). From the mobile phone segment, the test voice signal is fed into the second audio input on PC#1 recording the test file for VQ processing.

The background traffic generation setup provides background voice and data traffic over the AP#1 and AP#2. In order to control co-channel interference, each of AP together with PC is housed in one RF isolated chamber. AP is via internal antenna connected to PC.

The wired backbone LAN of the test setup is applied for the management of the test setup and execution of the preprogrammed test scenarios. The voice signal playback, recording and VQT processing – as well as the background traffic generation – are brought under control of a programmed test scenario, whereas the RF attenuation segment provided only for the manual variation and setting of RF signal attenuation to start the handover.

4 Test-bed parameters and test scenarios

Various test scenarios are developed in order to provide a specification and execution of VQ tests over a range of WLAN parameters and values of interest. The following selection of parameters is available for the test-bed configuration:

- **VoIP telephones**: due to no commercial phones available on the market, we can select from a notebook PC or fixed PC acting as a soft phone or a PC acting as a bridge between WLAN and a HW VoIP phone.
- **Signaling protocol**: because the hardware and software phones supporting SIP protocol work

better then phones supporting H323 protocol, chosen terminal equipment is determined to use SIP signaling protocol.

- **Background voice traffic**: Zero to 20-background voice channels are simulated for voice background traffic load. Each of these voice channels consists of two RTP streams with each of them encapsulating G.711 voice data. The RTP stream is captured in a file together with the timing data from actual VoIP conversation. This file is played by RTPSend utility producing a sequence of UDP RTP packets sent to remote computer and relayed back to the source computer thus forming a bi-directional voice channel.
- **Background burst traffic**: three levels of burst data traffic can be selected: no background burst data traffic, half channel capacity burst data traffic and saturated channel burst data traffic.
- **QoS**: we opt for two QoS setups in wireless and wired segments: without QoS and with QoS giving the highest priority to the UDP packets with destination ports numbered from 8000 to 8100. In wireless segment, Spectral Link Voice Priority protocol is enabled.
- **RF signal strength on AP for handover**: Two levels of RF signal strength is used, i.e. High and Low. High signal level is the highest possible level with no attenuation, while the Low signal level is set to a half of dynamic range of client receiver (-40dB) since there exist no possibility to monitor the signal of noise ratio at client's end.
- **Measurement repetitions**: In test procedures, we have several test loops that enable us to monitor the behavior of test equipment during a longer period of time. Measurements are performed on constantly active VoIP connection since we are unable to terminate/restart the test VoIP connection for each measurement. To minimize the influence of one measurement on subsequent measurements, a parameter was introduced so that a test voice file is played a specified number of times before actual VQ voice sampling started. Measurements are performed in sets of measurements under constant background traffic yet for each set of measurements background traffic is determined within background group loop and several background group loops are repeated in sequence loop in order to space evenly VQ measurements at the same conditions during a longer period of time.
- **Reference WAV file**: All VQ measurements are performed with 7.4 seconds WAV voice English spoken with short periods of silence reference file.

In order to automate execution of the VQ tests, the VQT scenario is implemented in the form of structured command files for batch execution, written in 4NT command interpreter.

All the tests are performed in two phases:

- a short test to select the best performing HW/SW; and
- a long test to measure the performance of selected HW/SW in detail.

Where multiple options are available, better performing HW or SW are selected so that the best available VoIP technology to WINDECT may be compared.

Each VQ test consisted of a number of steps:

- select and configure devices under testing,
- select RF signal strength parameters and (burst) data background traffic load,
- generate selected number of background VoIP streams (one pair of bidirectional RTP streams with G711 encoded voice data in it for each simulated VoIP channel),
- play the reference voice signal and record the test voice signal synchronously with the reference signal into a stereo WAV file repeat for averaging of results,
- process WAV file with Opera software and store VQT measures, and
- present VQT results.

The assessment of quality of a voice connection across a given WLAN/LAN test setup over a range of background voice and burst-type data traffic loading as well as over a number of other parameters determining the configuration and the execution of an actual test suite is bound to involve a large number (several hundreds or even thousands) of individual VQ measurements. The scope of testing is enlarged by an order of magnitude because of a need to statistically accommodate the spreads of measurement values due to stochastic nature of data traffic parameters as well as due to a lack of control over the distributed execution of test scenarios in a distributed Windows-based test setup. Credible test scenarios producing repeatable results could only be carried out in a relatively stable research laboratory and under – as much as possible – computer control of the test setup scenario execution.

5 WINDECT system performance evaluation

Voice quality measures PESQ, MOS and *R* are determined from the Opera tool. Ethereal software package is used for the packet loss measurement during tests involving handover. These measurements are performed so that the entire traffic to both APs are recorded by Ethereal connected on router port with mirrored traffic to both APs and after successful handover, SIP signaling and statistics of RTP streams are extracted and plotted using Ethereal statistics SIP extensions.

We execute several sets of voice-quality measurement test scenarios under different variations of test parameters. PESQ results for 50 repetitions of each measurement are presented in Fig. 3. Each set of 50 repetitions is executed under high burst traffic, QoS enabled and 0 to 20 background voice streams incremented in steps by 1. As it is expected the VQ is decreased by increasing a number of background voice channels. MOS drops for one degree from good quality described as perceptible but not unsettling the fair quality by increasing background voice channel from 1 to 20. Similar effect is observed studying average packed delay which is increased from acceptable 200ms up to even 1400ms.

Compared to other published 802.11b VoWLAN tests, we exhibit slightly better results as far as VQ is concerned. VQ drops significantly only when background data traffic increases close to channel saturation. In addition to this, a relatively small difference was observed in behavior with and without QoS. We attribute this to the fact that we (due to a lack of proper hardware test generators) simulated background traffic with a single stream of data. In a typical enterprise network, there are many streams generated by many devices communicating on network. Without QoS being activated, typical switch or router would try to forward these streams in the best effort manner (i.e. round robin). With only two streams, each of them still gets serviced frequently enough so that a moderate VQ is achieved even if prioritization is not enabled. If many background streams were available, a difference in delays and jitter between enabled and disabled QoS was significantly higher.

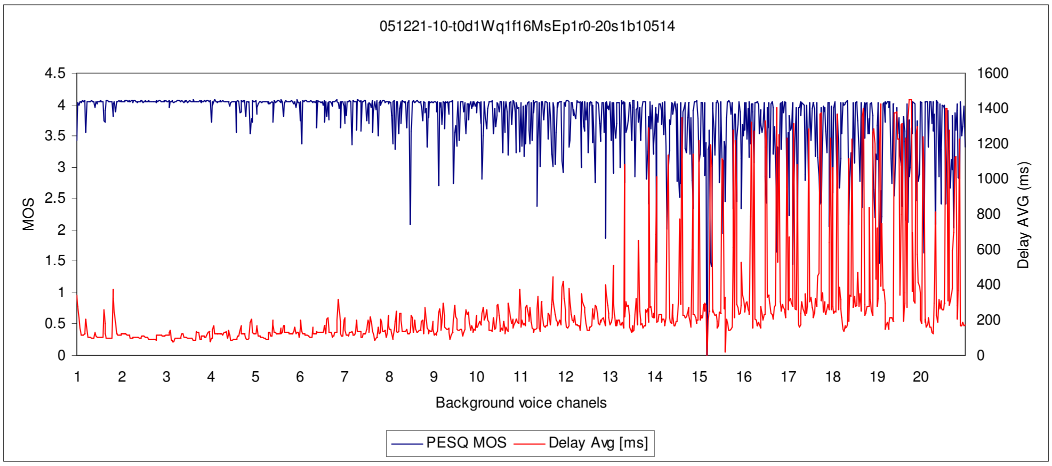

Fig. 3: PESQ results for 50 repetitions of each measurement (0 to 20 background voice streams, high burst traffic and QoS enabled)

In our measurements, by and large relatively good VQ measurements may be observed. With degradation of communication link between the VoIP phones, one can observe gradual degradation of average PESQ and at the same time larger spread of PESQ results. OoS (with single background stream) has limited effect and just lowers the spread of PESQ results.

Delayed and single missing packets have little or no influence on VQ, and VQ drops fast when there is no free high priority bandwidth left or with an increase in packet error rate due to a low signal/noise ratio on WLAN.

Voice signal delay increases with interference (delayed or missing RTP packets); it increases stepwise after an end of detected silence period, decreases gradually when there is error free RTP transfer; it slowly periodically increases and decreases when available bandwidth decreases close to the minimum required bandwidth (due to asynchronous clocks in both VoIP devices); frequency of these oscillations is proportional to channel quality, amplitude is inversely proportional to channel quality; it decreases sharply (on average) together with VQ when the remaining bandwidth drops close to or below required VoIP channel bandwidth (since terminal buffers cannot fill up).

We attributed the effect of oscillation (in a jigsaw form) in VQ measurement in saturated bandwidth conditions to the behavior of buffer length control algorithms embedded into the two IP phones of the test setup. The lengths of the buffer on one side of the VoIP connection varies depending on the loading of the VoIP channel and is based on information received over the RTP protocol from the other VoIP phone. With a degraded communications channel, this feedback becomes slower and affects first the observable delays and then VQ.

6 Conclusion

An experimental test-bed for objective ITU P.862 standard based VQT was constructed within a relatively small budget and absence of commercially available WLAN 802.11a–enabled mobile phones. Nevertheless, in the absence of comparable measurements of WLAN 802.11a-enabled equipment in academia, reasonably controlled RF environment, automating of testing scenarios and improved synchronization of captured voice samples with several thousand statistically analyzed measurement, results assured adequate qualification of experimental VQ test-bed.

ACKNOWLEDGEMENTS

The authors acknowledge the contribution of all partners of the WINDECT project. WINDECT is partially founded by the European Commission, the Swiss Federal Office for Education and Science, Slovenian Ministry of Higher Education, Science and Technology and Prevent Group, Slovenj Gradec.

References

[1] Converging enterprise Networks, White paper http://www.nortel.com/solutions/pt/es/collateral/nn102460-110602.pdf

[2] WINDECT, http//www.windect.ethz.ch

[3] ITU-T Recommendation P.800.1, *Mean opinion score (MOS) terminology*

[4] ITU-T Recommendation P.861, *Objective Quality measurement of telephone-band (300 - 3400 Hz) speech CODECs, describing an objective PSQM method for the estimation of the subjective quality from the objective measurement results*

[5] ITU-T Recommendation P.862, *PESQ an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech CODECs, describing an objective method for predicting the subjective quality of 3.1 kHz (narrowband) handset telephony and narrow-band speech CODECs*

[6] ITU-T Recommendation G.107, *The E-model, a computational model for use in transmission planning,* May 2000

[7] Opera Test Suite, http://www.opticom

[8] Agilent's Telephony Network Analyzers, http://www.agilent.com

[9] Powerful Multi-Platform Analysis Package, http://www.ethereal.com

[10] LAN switch/router Entrasys SSR2000, http://www.entrasys.com

[11] WLAN access point Entrasys AP3000, http://www.entrasys.com

[12] Zultys hardware phones models ZIP 4x5, http://www.zultys.com

[13] Real time protocol, RTP tools, http://www.cs.columbia.edu/IRT/software/rtptools

[14] D-ITG distributed traffic generator, (http://www.grid.unina.it/software/ITG/)

[15] 4NT command line interpreter by JP Software, http://www.jpsoft.com/