Real-time voice over wireless IP networks: The last challenge?

Ian Marsh
Laboratory for Computer Networks
Institute of Microelectronics and Information Technology
Royal Institute of Technology
Stockholm, Sweden
Email: ianm@sics.se
Advisor: Gunnar Karlsson

***Abstract*— Wireless connectivity is needed to bring IP-based telephony into serious competition with the cellular infrastructure. Nevertheless, quality problems remain with wireless VoIP services, not least when used with unlicensed spectrum technologies. This submission summaries on-going work in the area of real-time voice quality issues using a combination of networking and real-user assessment techniques.**

I. Introduction

The first part of this PhD work addressed problems of real-time voice issues over fixed IP networks [1]. The conclusion was that for communication paths using sufficiently dimensioned links and carefully designed end-points the 'system' could provide adequate interactive voice quality for the end-user. This fact has been successfully exploited by companies such as Skype Technologies S.A.

The remaining challenge is to ascertain whether wireless networks can also provide sufficient quality for its users. The availability of cheap IEEE 802.11 devices has increased the number of wireless access areas, particularly where wireless telephony is used; homes, offices and public spaces. Hence, there is new impetus for IP telephony deployment. Naturally, wireless access using unlicensed spectrum can suffer from severe disruptions caused by cheap electronic components, interference from base stations, other wireless devices, obstacles between the transmitter and receiver, other users traffic in the same cell and so on. These, and other factors, can deteriorate the quality of voice conversations to well below than that commonly accepted by consumers.

User satisfaction is important in the wide-scale acceptance and adoption of IP-based telephony services. IP networking researchers, thus far, have often made the mistake of ignoring the users' requirements where quality issues for real-time services are concerned. Widely adopted standards and quality tests are still defined by bodies such as the International Telecom Union (ITU).

Section II outlines some of the lower-layer factors affecting voice quality in wireless networks, such as the medium access behavior and environmental issues. Section III aims to show how we will combine this information obtained with the higher level aspects of real-time audiovisual communication, by involving more the *real* end-systems, people.

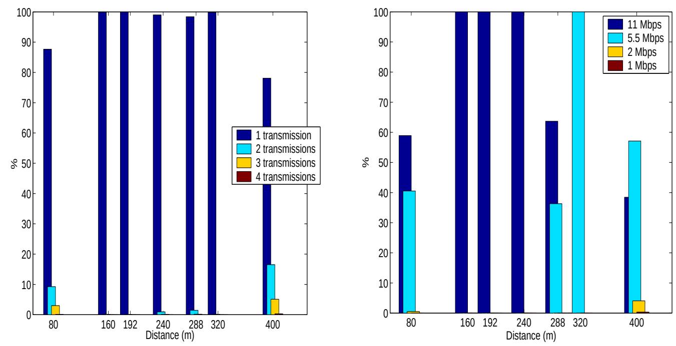

Fig. 1. Number of successful MAC transmissions and rates used vs. distance

II. Wireless voice quality measurements

We have conducted many hundreds of experiments and gathered several thousands of measurements in a variety of circumstances to assess the suitability of 802.11b networks for real-time voice communication. As an example, Figure 1 shows the effect of the environment between two nodes, outside, in ad-hoc mode, within line of sight of each other. The left plot shows a histogram of the number of transmissions at a particular distance, whilst the right plot shows the rate with which these transmissions took place. As a preliminary result, we can suggest that if a frame is lost, it may be beneficial to reduce the rate immediately. Also even at 80 meters interference was experienced, highlighting the interference problem.

We have found that the MAC layer behavior in *conjunction* with the application layer performance is both useful and informative in estimating the quality of VoIP sessions. For example, by recording the number of MAC retransmissions, one can predict if the terminal is entering a period of poor quality, so codec changes giving better loss robustness can be initiated, particularly if losses are the cause of the poor quality. Predication of impending poor quality by looking at IEEE 802.11 specific capabilities, such as signal-to-noise ratios, transmission rates, number of transmissions is a current topic within this research. The role of the environment is also an important factor for the loss characteristics. Intervening walls, windows, attenuate the signal causing sporadic losses, hence lessening the quality. As the speaker moves, this makes it that much harder to predict the upcoming conditions, more details on this topic can be found in [2].

III. End user quality assessment

Essentially two techniques exist for assessing the quality of speech, on-line methods such as the E-model, and off-line methods such as PESQ, both standardized by the ITU. In the case of on-line quality assessment, the E-model is the most suitable choice as its calculation can be done per packet. Table I shows the R-factor and the corresponding subjective ratings we are familiar with.

TABLE I

ITU's R-model and user satisfaction ratings

R-factor	User satisfaction
90	Very satisfied
80	Satisfied
70	Some users dissatisfied
60	Many users dissatisfied
50	Nearly all users dissatisfied

The R-factor value has been shown to be comprised of the following elements:

$$R = R_o - I_s - I_d - I_{e-eff} + A$$

R = rating value
R_o = basic SNR value (noise sources)
I_s = voice impairments (side-tones and quantization distortion)
I_d = delay and equipment impairments
I_{e-eff} = low bit-rate codecs (also random packet losses)
A = advantage factor (users accepting certain factors, e.g. they are only using a mobile)

Therefore the elements can be calculated separately and summed giving an R-factor that is almost *linear* with the real subjective ratings [3]. An important facet of the ITU's work is these quantities were chosen so that they (and their sub-components) can be simply summed as stated. In the equation above the delay is represented in the I_d component and the loss in the I_{e-eff} component. The values in fact, represent compound variables, however we omit their discussion due to space considerations. We have shown in Figure 2, the R-factors as functions of the delay for four different loss values of a G.711 codec, a standard PCM 64kbit/s speech coder.

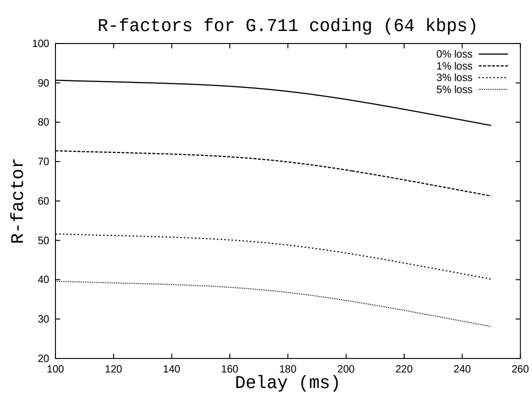

Fig. 2. R-factors for 100-250ms delays and 1-5% losses

The important point for our research is that under zero loss conditions this codec is considered ideal, as it does not distort the signal in any way, however as can be seen (the much lower R-factors for the loss values), losses severely affect its performance. Therefore, depending on the conditions, real-time decisions have to be made in order to maximize the R-factor at any instant. We are currently implementing R-factor approaches to optimizing quality in real-time voice systems.

However, the drawback of this E-model type approach is that depending on which parts of speech are lost (vowels, consonants, the start/end of a sentence etc.) varying quality ratings can be obtained for the same speech pattern. Therefore off-line methods such as the Perceptual Evaluation of Speech Quality (PESQ) have been developed by the ITU to estimate the speech quality based on a psycho-acoustic model [4]. The human quality effect is obtained by comparing the degraded speech sample with its clean version in the perceptual domain. PESQ is both time consuming in setting up and computationally, as it involves using (and comparing) standard samples of different sexes, ages and languages, so it cannot be used in real-time, however we will use it as an off-line quality metric for wireless VoIP sessions.

IV. Conclusions

We believe we have struck a good balance between networking research and user-assessment techniques in order help advance real-time services using wireless IP networks. This includes looking more closely at the lower layers as well as the important user aspects as briefly exposed in this abstract. We also want to leverage the results from the standardization committees into VoIP services. Finally as well as advancing our own doctoral work, we belive that our measurement data and expertise is useful to the networking community as a whole.

Acknowledgments

Thanks to Professors Gunnar Karlsson and Gerald Maguire at IMIT, KTH, (graduated) masters' students Juan Carlos Martín Severiano and Victor Yuri Diogo Nunes plus Björn Grönvall the 'master mentor' at SICS.

References

[1] I. Marsh, "Quality aspects of audio communication (Licentiate Thesis)," May 2003, TRITA-IMIT-LCN AVH 03:01, ISSN 1651-4106, ISRN KTH/IMIT/LCN/AVH-03/01 SE.

[2] J. C. M. Severiano, G. Maguire, I. Marsh, and V. Y. D. Nunes, "Cross-layer measurements using voice in 802.11b networks," in *2nd Swedish National Networking Workshop*, Nov. 2004.

[3] International Telecommunication Union, "The E-model, a computational model for use in transmission planning," Recommendation G.107, Dec. 1998.

[4] International Telecommunication Union, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," ITU-T Recommendation P.862, Feb. 2001.