7D　プロトコル

座長　重野　寛（慶大）

「マルチメディア，分散，協調とモバイル
（DICOMO 2006）シンポジウム」平成 18 年 7 月

Performance Analysis of VoIP over WLAN

Razvan Beuran [1,2] Yoichi Shinoda [2,1] Shin-ichi Nakagawa [1,2]
Junya Nakata [2,1] Toshiyuki Miyachi [2,1] Ken-ichi Chinen [2,1] Yasuo Tan [2,1]
[1] National Institute of Information and Communications Technology
[2] Japan Advanced Institute of Science and Technology
razvan@nict.go.jp

1. INTRODUCTION

Recent disasters such as the Indian Ocean tsunami in 2004, or the 2005 hurricanes Katrina and Rita in U.S.A. have shown the importance of emergency communication systems. This is true both in the preceding phase of the events, to issue warnings and evacuation instructions, as well as during and after catastrophes happen, to coordinate the activity of rescue teams. Dependable communication is also essential in mission-critical and safety-critical systems, or even in normal business environments that require "anytime, anywhere" access to network resources.

Unfortunately these recent events have also demonstrated that current communication systems fail too easily under emergency conditions. An alternative is to make a more extensive use of wireless networks (WLANs). Using converged WLANs one may transmit both audio and video information, so that rescuers can communicate with each other and with remote experts, as well as receive data, such as street maps or building floor plans – all crucial for saving lives and preventing losses.

Wireless LANs are more stable than other communication infrastructures, as they are decentralized. Since their potential failure is independent, *ad-hoc* WLANs can continue functioning in emergency conditions. The nodes of WLANs are cheap and require little power. Moreover, the potential of using advanced features on WLANs that are not available in traditional communication systems makes it possible to provide probabilistic guarantees of service for emergency responders. Using a priority-enforcement system such users could be given an assured service level, independently of the activity of regular users.

In this paper we present our approach to investigating the possibility of using real-time applications, such as Voice over IP (VoIP), on WLANs in a dependable manner. According to the survey we did in [1], this is difficult at the moment for several reasons:

i. WLAN QoS parameters (bandwidth, packet loss, delay & jitter) have a high variability in real-world environments, and this has a significant effect on application performance;
ii. Existing WLAN QoS mechanisms are only of limited use for managing contention when applications with different QoS requirements, such as VoIP calls and TCP-based data traffic, share the same communication channel;
iii. VoIP is a multimedia application that requires timely servicing of the voice traffic; this is a challenging task in WLANs, even when using QoS enforcement, since most currently-implemented QoS mechanisms focus on bandwidth provisioning;
iv. Roaming between access points, a typical WLAN event, introduces communication gaps that may even be of the order of seconds, an unacceptable situation for real-time applications.

We started analysing application performance over WLAN, and VoIP in particular, using StarBED [2], the large scale network experiment environment of Hokuriku Research Center in Ishikawa, Japan. StarBED is a cluster-based testbed currently employing about 700 PCs. The custom-designed configuration language SpringOS makes it possible to define complex experiments on StarBED in a straightforward manner.

The paper is structured as follows. First we present the analysis methodology that we propose for the study of application performance over wireless LANs, and emphasize the specific issues related to VoIP. Then we discuss the use of WLAN emulation, which is a key element of our approach as a complement to real-world tests. Following that we show some illustrative results for our study of VoIP performance on WLAN. The paper ends with a section of conclusions and future work.

2. ANALYSIS METHODOLOGY

The first step to take in studying IP application performance and dependability on WLANs is to define a test methodology that allows to objectively assess application performance and understand its dependence on network conditions. The methodology we propose is application independent. We use it here to study VoIP as a representative real-time application, since voice communication is one of the most important forms of communication in emergency situations.

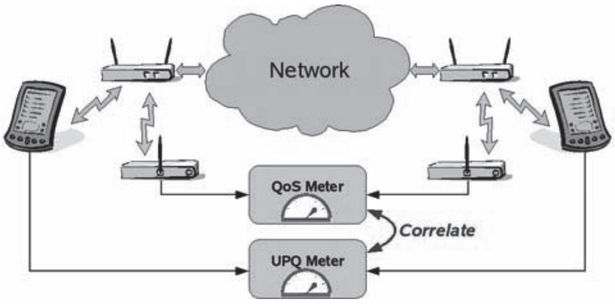

Figure 1: Experimental setup for application performance assessment on WLAN.

The setup in Figure 1 is adapted from the system we previously used to study VoIP performance on wired networks [3]. Using this setup we carry out a two-level analysis. At the level of the network we investigate the performance issues of WLANs, such as the dependency on signal strength and number of access points, quality degradation management techniques, etc. This is done objectively through the use of the "QoS Meter" block that computes the WLAN QoS parameters based on the monitored traffic traces. Simultaneously, at application level, we measure the User-Perceived Quality (UPQ) for the application under study. In the case of VoIP we use methods such as the ITU-T recommendations G.107 [4] and P.862 [5]. This function is performed by the "UPQ Meter" block.

Correlating the network and application performance permits us to establish objectively what are the requirements of an application such as VoIP in order to ensure user satisfaction, as well as determine what type of mechanisms are needed to meet these requirements on WLANs.

One important aspect of this setup is its ability to capture the dynamic behaviour of the tested systems. Running averages and global assessments are not meaningful for short-term performance issues, which are nevertheless critical in disaster situations. Moreover WLANs are dynamic environments by definition: signal reception conditions fluctuate, the number of nodes and their position vary, the access points with which the nodes communicate, or their peers in *ad-hoc* networks change. By capturing the dynamic behaviour of the network we can follow application performance fluctuations over time.

The key element for the aforementioned task is the use of WLAN traffic monitors (sniffers). They allow access to physical-level WLAN information, such as signal strength, which is invaluable for understanding the performance issues, and difficult to obtain otherwise from devices such as wireless phones or Personal Digital Assistants (PDAs). Another important feature is the accurate timing of the events and the time synchronization between the monitors in different points of the network.

As mentioned before our goal is to objectively establish the relationship between physical environment, WLAN level performance, and in the end user satisfaction. For the physical environment metrics such as the power of the received signal (*Pr*) are used. At network level we measure the three inter-dependent QoS parameters: bandwidth, packet loss, and delay & jitter. In order to estimate the User-Perceived Quality application specific metrics must be employed. In the case of VoIP two methods are more appropriate: the R-value score, which is the result of the E-model described in the ITU-T Recommendation G.107 [4], and the PESQ score, proposed by the ITU-T Recommendation P.862 [5]. Table 1 shows a concise comparison between the two metrics.

The R-value can be converted to the Mean Opinion Score (MOS) scale [6]. A value of 4.5 on the MOS scale indicates optimum quality, with good quality being associated to scores higher than 3. Quality is considered acceptable for scores between 2 and 3, whereas scores lower than 2 indicate unacceptable quality.

R-value	PESQ score
Requires traffic measurements	Requires voice recording
Test pre-determined conditions only	Independent on conditions
Easy to compute, but parameter values are key	Uses complex psycho-acoustic models
Accounts for one-way delay	Listening-only score
Jitter needs separate treatment	Jitter is taken into account

Table 1: R-value versus PESQ score.

Most current studies use the R-value to predict VoIP UPQ. Given the way in which it is computed and the fact that the E-model is intended for transmission planning, we consider it suitable mainly for making preliminary predictions of the VoIP UPQ. Our approach makes it possible to capture voice data when performing an experiment. Therefore we also compute the PESQ score, which is an objective estimate of VoIP UPQ. The value of the PESQ score reflects more accurately the effective network degradation that took place at the moment of the experiment, since it makes use of the voice signals at the input and at the output of the communication channel. See section 4 for a discussion of results obtained with these two metrics.

3. WLAN EMULATION

Real-world experiments are only one aspect of our research. Such experiments are very useful for understanding the behaviour of real WLAN systems. However the range of conditions that can be tested in real-world experiments is limited and difficult to control.

WLAN emulation is an integral part of our approach to studying application performance on wireless LANs. Emulating the WLAN environment enables us to study a wide range of controllable network conditions. Emulation uses real applications, and the setup we use in this case (see Figure 2) is similar to that presented in section 2.

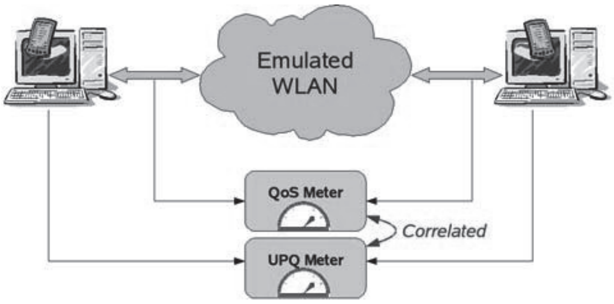

Figure 2: WLAN emulation for application performance assessment.

In Figure 2 the applications that run on the end PCs are of the same type with those running on a PDA, or a wireless phone. Therefore using this setup we can test the performance of the same applications as those in the real-world experiments. Moreover, in this setup too we measure the QoS parameters and the UPQ. Correlating these two measures makes it possible to determine objectively the relationship that exists between application performance and network quality degradation under varying network conditions.

There are several key requirements for such a methodology to be effective, and our approach addresses them. Realism of the emulation is important because it allows drawing conclusions that are thereafter useful in real-life deployment. Most application performance studies focus on bandwidth-oriented applications, typically based on TCP/IP. Models used are oversimplified in a bandwidth-oriented perspective, and do not take into account those aspects that are most important for real-time applications, such as the delay variation and packet loss, for example those associated with rate changes.

The transition from physical layer network effects, which are consequences of events in the physical world (signal strength variation, node movement etc.), to data-link layer and network layer effects is a task that requires the use of several techniques: modelling of 802.11 protocols and understanding their behaviour, analysing real implementations of 802.11 protocols, and finally reproducing the interaction with the application traffic, followed by the objective assessment of UPQ.

The emulation approach we use transforms a user-meaningful real-world representation of a WLAN environment into a network quality degradation description, termed "ΔQ description". The ΔQ description is sufficient to subsequently configure a network emulator such as *dummynet* [7] and effectively reproduce an environment that corresponding accurately at network level to the emulated scenario. We ran in this environment a VoIP application and studied its performance as described next.

4. EXPERIMENTAL RESULTS

In this section we show some illustrative results for VoIP performance over WLAN. These results were obtained using the setup in Figure 2. Consider a scenario with a node that from an initial distance of 10 m with respect to an access point moves on a perpendicular direction with a speed of 0.5 m/s for a time interval of 30 s. This is a scenario fragment representative for a user moving in a building while making a VoIP over WLAN phone call.

Using a log-distance path loss model [8] with α = 8 (difficult reception conditions) and $Pr0$ = -20 dB we obtain the power of the received signal (Pr) shown in Figure 3 together with the distance between the access point and the mobile node. Horizontal dashed lines indicate the thresholds between minimum signal power levels for different operating rates (from top to bottom, 11 and 5.5 Mb/s respectively). Note that Pr falls under the 11 Mb/s threshold after $t \cong 20$ s and below the 5.5 Mb/s threshold at $t \cong 26$ s.

We modelled the WLAN effects of the signal power variation by taking into account realistically the consequences the Pr variation has on the quality degradation ΔQ. Using this ΔQ as the input of the ITU-T E-model we calculated the R-value, and predicted what the user satisfaction would be if communicating under such circumstances.

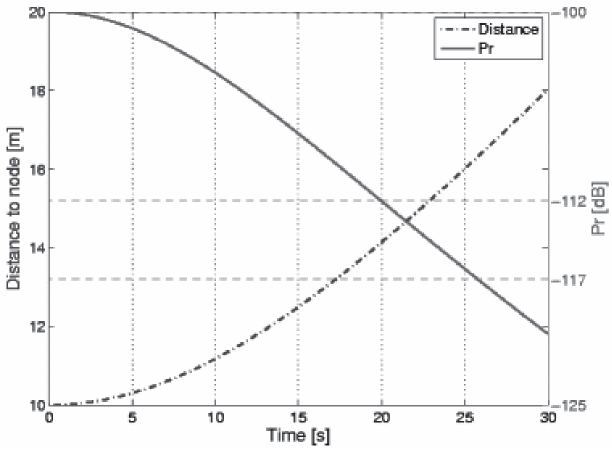

Figure 3: Distance between access point and mobile node, and the power of the received signal versus time.

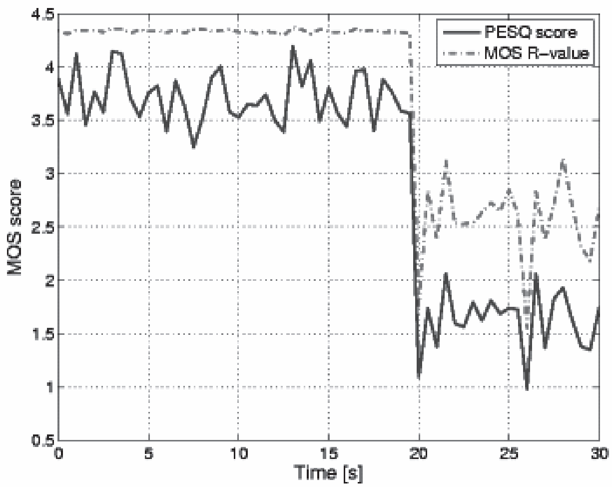

Figure 4: PESQ score and MOS R-value versus time.

The next step was to use the same ΔQ descriptors to drive the network emulator *dummynet* while real voice data was sent through the network. We used a customized version of the SpeakFreely 7.6a application that saves the output voice signal [9]. For this test SpeakFreely was configured to make use of the codec G.711 [10]. The voice input was one of the standard voice test files supplied with ITU-T P.862. Based on the input and output voice signals we computed the PESQ score. Both the MOS R-value (the

MOS score obtained analytically from the previously computed R-value) and the PESQ score are represented in Figure 4.

We can observe in Figure 4 that the MOS R-value has generally higher (i.e. more optimistic) values than the PESQ score, which is deemed to be more realistic. In the case when $t < 20$ s signal conditions are good, therefore the average packet loss was considered low (1%). Hence the MOS R-value is stable around a value of 4.3. However the PESQ score has a much more evident variation. The explanation consists in the fact that while average packet loss values over a long interval might be 1%, over shorter timescales and for a small number of packets the loss values change significantly, and this change leads to a important perceived quality variation. Note that while this is true in the range of low packet loss values, the variation has less significant effects for high packet loss values.

As remarked previously, at times $t \cong 20$ s and $t \cong 26$ s the low level of the received signal power triggers rate transitions, with obvious effects on VoIP UPQ. However these effects are not caused by the rate change in itself (which only goes down to 2 Mb/s in the worst case, largely sufficient for the roughly 80 kb/s stream generated by the G.711 codec). On the contrary, these effects are produced by the packet loss and delay variation that occur just before and during the rate transition, and that are ignored in most research papers. Moreover, once Pr falls under the optimum reception threshold for 11 Mb/s, the auto-rate fallback mechanism in 802.11 is triggered, and this causes additional quality degradation due to the packet loss induced when trying to send at higher rates.

5. CONCLUSION

Disaster situations impose stringent requirements on communication systems, and in such cases dependability and performance guarantees are essential. We started to investigate VoIP performance in WLAN environments, since speech is one of the most important forms of communication in emergency situations.

In a first stage we started using real-world tests and a WLAN emulated environment to determine the relationship between the events in the physical world associated with WLANs, the corresponding network quality degradation and in the end the user perceived quality for VoIP. Our methodology lays emphasis on dynamic behaviour capture and emulation realism.

We illustrated the practical use of our approach by emulating a real-world scenario, for which we determined the induced network quality degradation. Then we quantified the influence of this quality degradation on VoIP UPQ in an objective manner using ITU-T recommendations concerning expected user satisfaction for VoIP communication.

In a second phase of our research we will investigate application performance assurance. We intend to use the same setup to study techniques of application performance assurance in WLAN environments that could be used in emergency conditions, such as disaster rescue operations and other critical situations. The framework and the specific techniques that we plan to develop will allow the creation of *ad-hoc* WLANs under critical conditions, and the assurance of quality guarantees even under such circumstances for high-priority users (e.g., public safety teams, hospitals, etc.). The IEEE 802.11e standard for QoS on WLAN will be studied first to determine whether it is suited for use in the context of emergency communication systems.

REFERENCES

[1] R. Beuran, "VoIP over Wireless LAN Survey", research report, IS-RR-2006-005, Japan Advanced Institute of Science and Technology (JAIST), Ishikawa, Japan, April 2006.

[2] StarBED – A Large Scale Network Experiment Environment, http://www.starbed.org.

[3] R. Beuran, M. Ivanovici, "User-Perceived Quality Assessment for VoIP Applications", technical report (delivered to U4EA Technologies), CERN-OPEN-2004-007, CERN, Geneva, Switzerland, January 2004.

[4] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning", ITU-T, March 2005.

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

[6] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", ITU-T, August 1996.

[7] Dummynet, FreeBSD network emulator, http://info.iet.unipi.it/~luigi/ip_dummynet.

[8] T. S. Rappaport, "Wireless Communications – Principles and Practice", Prentice Hall PTR, 1996.

[9] B. C. Wiles, J. Walker, "Speak Freely 7.6a", http://www.speakfreely.org.

[10] ITU-T Recommendation G.711, "Pulse Code Modulation (PCM) of voice frequencies", ITU-T, 1993.