Wireless VoIP at Home: Are We There Yet?

Gerardo Rubino, Martín Varela and Jean–Marie Bonnin *

15th May 2005

Abstract

Over the last few years, and especially since the advent of wireless technologies such as IEEE 802.11, home networks are becoming ubiquitous. There is also an important increase in the number of broadband–connected homes, and many new services are being deployed by broadband providers, such as TV and VoIP.

The home network is thus becoming the "media hub" of the house. This trend is expected to continue, and to expand into the Consumer Electronics (CE) market as well. This means new devices that can tap into the network in order to get their data, such as wireless TV sets, gaming consoles, tablet PCs, etc. As more and more devices are networked, we may see a new bottleneck, namely the wireless LAN (WLAN). The most widely deployed wireless technologies for home networks are currently IEEE 802.11b and 802.11g, which provide nominal top bandwidths of 11 and 54Mbps respectively. To a lesser extent, there are also Bluetooth connected devices, which can work at about 2Mbps in good conditions. In practice, the throughputs observed in *wi–fi* networks range from 4–5Mbps for 802.11b to 20-25Mbps to 802.11g. Besides, the network performance varies depending on the topology of the network, environmental parameters, and interference from other equipment. These performances are low when compared to current Ethernet LANs (100Mbps – 1Gbps), but the lack of cabling makes wireless networking an attractive option for home use. As new devices, applications and services continue to be developed, it is very likely that the load imposed on the network will increase, shifting the bottleneck from the access network to the wireless links.

Besides the limited bandwidth available in *wi–fi*, the nature of wireless networks is such that their performance is very variable. Impairments due to the position of the hosts, multi-path propagation, hosts with weaker signal strength, etc. might drastically decrease the network's performance at any time. This in turn implies a strong variability in the perceived quality of real–time applications such as voice streams, which may have a big impact on the adoption of these new technologies.

In this paper we study the quality of VoIP flows in the context described above. To this end, we propose a stochastic model for the *wi–fi* network, which closely mimics its performance variations. In order to perform the quality assessments, we couple this model with a Neural Network–based quality assessment tool [1, 2], which provides MOS results that correlate very well with human perception.

1 Introduction

There is nowadays an increasing number of multimedia services being offered over the Internet, ranging from VoIP to video–conferencing. These have important requirements in terms of network QoS, which the Internet does not currently provide. The Internet infrastructure was not designed with these types of applications in mind, so traditionally, the quality of such services has not been optimal. Moreover, it can be said that most of the QoS–related problems are currently found in the access networks, and not in the core or in the local area networks (LANs). This is due to the fact that the core networks are usually over–dimensioned, and that LANs are usually very fast (100Mbps Ethernet LANs tend to be the norm for Small Office and Home, or SOHO, environments, and 1Gbps LANs are not rare these days)

However, the increasing number of wireless LAN (WLAN) deployments (normally in the form of IEEE 802.11b or 802.11g networks) coupled with a sharp increase in the number of homes with broadband connections, and new services being currently marketed might soon change that situation. As more and more devices are networked, we may see a new bottleneck, namely the

*IRISA/INRIA Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France. E-mails: {Gerardo.Rubino, Martin.Varela, Jean-Marie.Bonnin}@irisa.fr

WLAN. The most widely deployed wireless technologies for home networks are currently IEEE 802.11b and 802.11g, which provide nominal top bandwidths of 11 and 54Mbps respectively. To a lesser extent, there are also Bluetooth connected devices, which can work at about 2Mbps in good conditions (for Bluetooth 2.0). In practice, the throughputs observed in *wi–fi* networks range from 4–5Mbps for 802.11b to 20-25Mbps to 802.11g. Besides, the network performance varies depending on the different devices are connected, environmental parameters, and interference from other equipment. These performances are very low when compared to current Ethernet LANs, but the lack of cabling makes wireless networking an attractive option for home use.

As new devices, applications and services continue to be developed, it is very likely that the load imposed on the network will increase, shifting the bottleneck from the access network to the wireless links. As of this writing, there are already several European ISPs offering TV channels over DSL lines, and VoIP services, for what can be considered very cheap prices. It is reasonable to assume that this will further increase the adoption of broadband and of these services. The next logical step is to allow for wireless TV sets, music players, and the like to tap into the home network in order to get their data. Considering that a single TV–quality stream may require between 1.5 and 10Mbps of bandwidth, depending on encoding and quality parameters (and that's just TV, not HDTV), the 5 to 25Mbps (shared by all stations for down–link, up–link and local traffic) offered by the current crop of *wi–fi* networks suddenly seem a bit limiting. Of course, these networks will not be only used for multimedia traffic. More traditional uses such as E–mail, WWW, file transfers (notably P2P, which is quite demanding in terms of bandwidth both down and up–stream), gaming, etc. will continue to develop. It is also likely that nuisances like worms and viruses will also continue to exist, more so in an always–on networking environment where most users are not really careful, and they will too compete for the available resources.

Besides the limited bandwidth available in *wi–fi*, the nature of wireless networks is such that their performance is very variable. Impairments due to the position of the hosts, multi-path propagation, hosts with weaker signal strength, etc. might drastically decrease the network's performance at any time. This in turn implies a strong variability in the perceived quality of real–time applications such as voice streams, which may have a big impact on the adoption of these new technologies. In fact, as these services are not generally free, their perceived quality should be at least comparable to that of their traditional counterparts, which means at least *toll quality*[1] for VoIP. Recent results [3] show that VoIP performance over wireless networks is quite limited. In that paper, the authors used the network only for VoIP flows (which are not especially bandwidth–hungry), and showed that the quality was severely degraded even for a relatively low number of flows. When considering other applications which may be found in a media–centered home network and physical aspects such as the spatial distribution of the hosts, it is clear that VoIP quality will decrease. What we are interested in is just how much it will decrease, and whether the quality levels obtained will be acceptable.

The rest of the paper is organized as follows. In Section 2 we describe the most relevant characteristics of WLANs and present the network model we used for our simulations. Section 3 briefly describes the methodology used to measure the perceived quality, and presents the results obtained. Finally, we conclude and present future research directions in Section 4.

2 The Network Model

As mentioned in Section 1, we consider a home network context, where a high–speed xDSL or cable connection is available, and a wireless router provides access to it for different kinds of devices. Although in this study the actual speeds are not as relevant as the load of the network, we will use values comparable to those found in an 802.11b *wi–fi* network. We will assume that the link to the Internet provides enough bandwidth for all the flows considered (i.e. it is capable of outperforming the WLAN, which is the case for newer xDSL or cable connections, which provide speeds easily upwards of 20Mbps). In our model, several devices such as PCs, PDAs, gaming consoles, TVs, etc. are connected to the wireless network in order to access the Internet, and to have connectivity among them. Some of these devices run ”standard” TCP or UDP applications, such as file transfers, e–mail, P2P, etc. while others run real–time applications (in this context we use this term to refer to those applications that have strong temporal constraints), normally RTP over UDP traffic, such as VoIP, video on demand, video-conference, or multi-player gaming. We assume that most of these applications are working over the Internet, and therefore the bulk of the traffic has to pass through a unique *wi–fi* router / access point (AP) and share the radio channel.

The wireless network's performance is very variable. Factors such as signal–to–noise ratio variations (SNR) due to internal and external interferences (which in turn depend on the radio frequency used – 2.4GHz or 5GHz), signal fading, access contention mechanisms, and multi-path propagation all contribute to the degradation of the actual network's performance.

[1] That is, the quality provided by the PSTN.

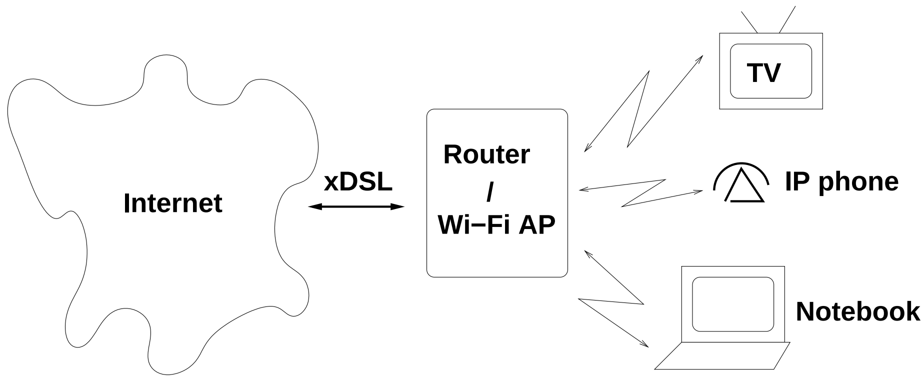

Figure 1: Simplified view of the network model. A xDSL connection relies the home network to the Internet, and a wireless router / access point connects the devices to the xDSL line, and among them.

Variations in the SNR often imply a need to adapt to a fluctuating radio channel (Link Adaptation Mechanism) by means of more robust coding. This comes at a price, namely a drop in the effective bandwidth. The standards allow for several functioning modes, each one resulting in a smaller effective bandwidth, so globally, the available bandwidth of a *wi–fi* network varies in a discrete fashion among several levels. For instance, for 802.11b, there are four operation modes [4], as shown in Table 1.

Mode	Capacity	Theoretical throughput
1 Mbps	0.93	0.93 Mbps
2 Mbps	0.86	1.72 Mbps
5.5 Mbps	0.73	4.02 Mbps
11 Mbps	0.63	6.38 Mbps

Table 1: 802.11b operation modes and maximal theoretical speeds.

When too many losses are being detected, or when the SNR is too low, the network will go into the next lower mode. This behavior makes for violent variations in the available bandwidth, which greatly impacts the performance of real–time applications. Other factors can also degrade network performance. For example, if a host is having a weak signal, it will use a low–bandwidth mode, in turn decreasing the effective bandwidth for the other hosts [5]. This happens because the contention avoidance mechanism will give all stations the same probability of access to the radio channel, and stations with a lower rate mode need more time to send the same amount of data. The transmission can also experience short loss bursts [6] due to temporary radio perturbation and delay spikes due to contention at the MAC level.

For our experiments, we used a simplified network model. We represent the network bottleneck (the AP in this case) by means of a multi-class ./M(t)/1/H queue, in which the server capacity varies with time in a way similar to that of a *wi–fi* network. The server capacity hops among four different levels, the time spent on each one being an exponentially distributed random variable. After that, it either goes up (or stays at the maximum) or down (or stays at the minimum) one level. Our traffic is split in two classes, background and real–time. Our background traffic consists of file transfers (web or FTP), P2P applications, e–mail, etc. It is mostly composed of TCP traffic. This is modeled using an ON/OFF process, which allows to control the burstiness of the flow. The real–time traffic is an aggregate of audio, video, voice, game streams, etc. This kind of traffic is less bursty, so we modeled it as Poisson process. We considered that real–time traffic accounted for 50% of the packets transmitted.

As in [7], and based on the data used in [8], we chose an average packet size of about 600B, which we used to calculate the different service rates of the router/AP. The queue size of the router/AP was set to 250 packets, and the scheduling policy used was FIFO. In order to keep the model as simple as possible, we did not represent the short loss spikes mentioned above. This should not be a problem, since at the load levels we considered, most of the losses at the IP level will be due to congestion at the router/AP, and will thus be significantly higher than those caused by contention at the MAC layer.

Figure 2 shows a bandwidth trace for a 200s period using the model described above. In Figure 3 we can see the real–time traffic loss rates (averaged every 5s) for the same period and for a medium load case (about 800 pps). It is easy to see how the loss rate

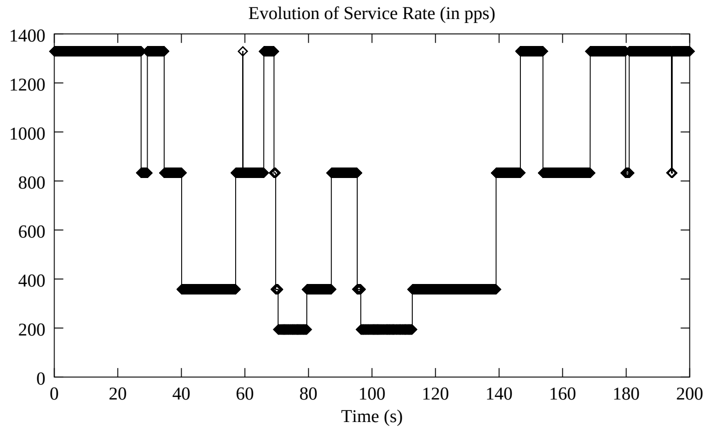

Figure 2: Sample 200s bandwidth trace. The mean throughput was set to 800pps for this simulation.

varies with the available bandwidth. Normally, voice streams are quite resilient to network losses, but with values as high as the ones obtained, there will be a very noticeable drop in the perceived quality. Furthermore, the loss process is such that relatively long bursts are common. In Figure 4, we can observe mean loss burst size (MLBS) values as high as 5, which have a clearly negative impact on the effectiveness of FEC, if it is being used.

3 Quality Assessment Results

In this Section we present the VoIP quality results we obtained for our scenario. The method which we used to quantify the quality of the voice stream as perceived by the end–user is called Pseudo–Subjective Quality Assessment (PSQA), and it is presented at large in [9, 1]. We will briefly describe it in the following subsection.

3.1 PSQA

The method we used is a hybrid between subjective and objective evaluation called Pseudo–Subjective Quality Assessment (PSQA). The idea is to have several distorted samples evaluated subjectively, and then use the results of this evaluation to teach a Random Neural Network (RNN) the relation between the parameters that cause the distortion and the perceived quality. In order for it to work, we need to consider a set of P parameters (selected *a priori*) which may have an effect on the perceived quality. For example, we can select the codec used, the packet loss rate of the network, the end–to–end delay and/or jitter, etc. Let this set be $\mathcal{P} = \{\pi_1, \ldots, \pi_P\}$. Once these *quality–affecting* parameters are defined, we choose a set of representative values for each π_i, with minimal value $\pi_{\min}$ and maximal value $\pi_{\max}$, according to the conditions under which we expect the system to work. Let $\{p_{i1}, \cdots, p_{iH_i}\}$ be this set of values, with $\pi_{\min} = p_{i1}$ and $\pi_{\max} = p_{iH_i}$. The number of values to choose for each parameter depends on the size of the chosen interval, and on the desired precision. For example, if we consider the packet loss rate as one of the parameters, and if we expect its values to range mainly from 0% to 5%, we could use 0, 1, 2, 5 and perhaps also 10% as the selected values. In this context, we call *configuration* a set with the form $\gamma = \{v_1, \ldots, v_P\}$, where v_i is one of the chosen values for p_i.

The total number of possible configurations (that is, the number $\prod_{i=1}^{P} H_i$) is usually very large. For this reason, the next step is to select a subset of the possible configurations to be subjectively evaluated. This selection may be done randomly, but it is important to statistically cover the points near the boundaries of the configuration space. It is also advisable not to use a uniform distribution,

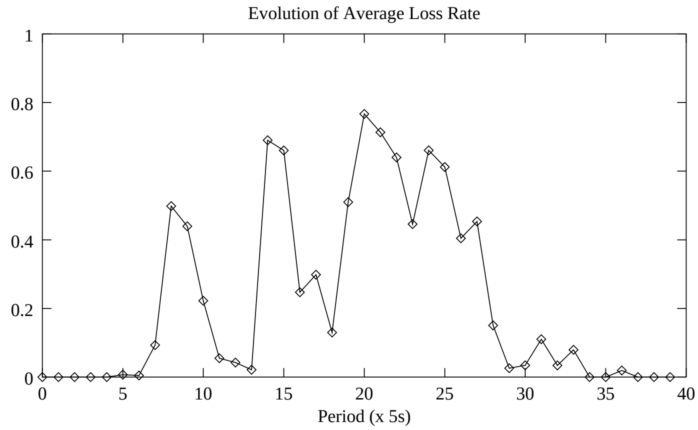

Figure 3: 5–second average loss rates (real–time traffic) for the same 200s period.

but to sample more points in the regions near the configurations which are most likely to happen during normal use. Once the configurations have been chosen, we need to generate a set of "distorted samples", that is, samples resulting from the transmission of the original media over the network under the different configurations. For this, we use a testbed, or network simulator, or a combination of both.

We must now select a set of M media samples (σ_m), $m = 1, \cdots, M$, for instance, M short pieces of audio (subjective testing standards advise to use sequences having an average length of 10s). We also need a set of S configurations denoted by $\{\gamma_1, \cdots, \gamma_S\}$ where $\gamma_s = (v_{s1}, \cdots, v_{sP})$, v_{sp} being the value of parameter π_p in configuration γ_s. From each sample σ_i, we build a set $\{\sigma_{i1}, \cdots, \sigma_{iS}\}$ of samples that have encountered varied conditions when transmitted over the network. That is, sequence σ_{is} is the sequence that arrived at the receiver when the sender sent σ_i through the source-network system where the P chosen parameters had the values of configuration γ_s.

Once the distorted samples are generated, a subjective test [10, 11] is carried out on each received piece σ_{is}. After statistical processing of the answers (mainly for detecting and eliminating bad observers, that is, observers whose answers are not statistically coherent with the majority), the sequence σ_{is} receives the value μ_{is} (often, this is a *Mean Opinion Score*, or MOS). The idea is then to associate each configuration γ_s with the value

$$\mu_s = \frac{1}{M} \sum_{m=1}^{M} \mu_{ms}.$$

At this step, there is a quality value μ_s associated with each configuration γ_s. We now randomly choose S_1 configurations among the S available. These, together with their values, constitute the "Training Database". The remaining $S_2 = S - S_1$ configurations and their respective values constitute the "Validation Database", reserved for further (and critical) use in the last step of the process.

The next phase in the process is to train a statistical learning tool (in our case, a RNN) to learn the mapping between configurations and values as defined by the Training Database. Assume that the selected parameters have values scaled into [0,1] and the same with quality. Once the tool "captured" the mapping, that is, once the tool trained, we have a function $f()$ from $[0,1]^P$ into $[0,1]$ mapping now any possible value of the (scaled) parameters into the (also scaled) quality metric. The last step is the validation phase: we compare the value given by $f()$ at the point corresponding to each configuration γ_s in the Validation Database to μ_s; if they are close enough for all of them, the RNN is validated (in Neural Network Theory, we say that the tool *generalizes well*). If the RNN did not validate, it would be necessary to review the chosen architecture and configurations. In the studies we have conducted while developing this approach, not once did the RNN fail to be validated. In fact, the results produced by the RNN are generally

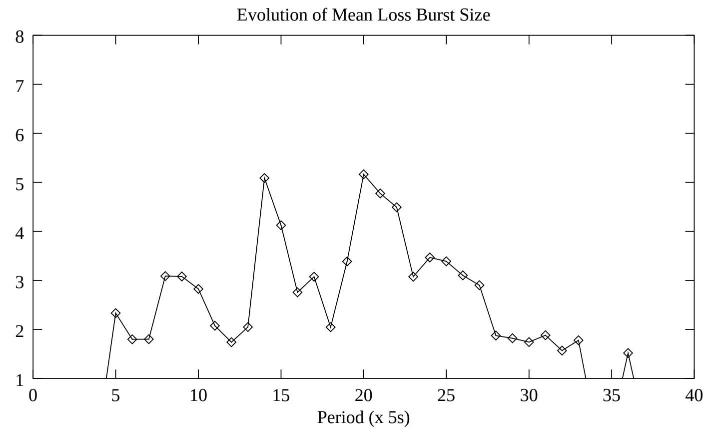

Figure 4: Mean loss burst sizes (sampled every 5 seconds for real–time traffic) for the same 200s period.

closer to the MOS than that of the human subjects (that is, the error is less than the average deviation between human evaluations). As the RNN generalizes well, it suffices to train it with a small (but well chosen) part of the configuration space, and it will be able to produce good assessments for any configuration in that space.

3.2 Quality Assessment Results

Let us now discuss the results we obtained for our wireless home network scenario. As seen in Figure 2, the available bandwidth is very variable with time. This, given a moderately high load in the network, results in very important loss spikes, as shown in Figure 3. While VoIP streams can maintain relatively good quality levels in the event of mildly high loss rates, the ones that happen in this kind of context are much too high. Figure 5 shows the PSQA scores for the same 200s period considered in the previous figures. It can be seen that the perceived quality is well below acceptable levels (a score of 3 is considered acceptable in this scale).

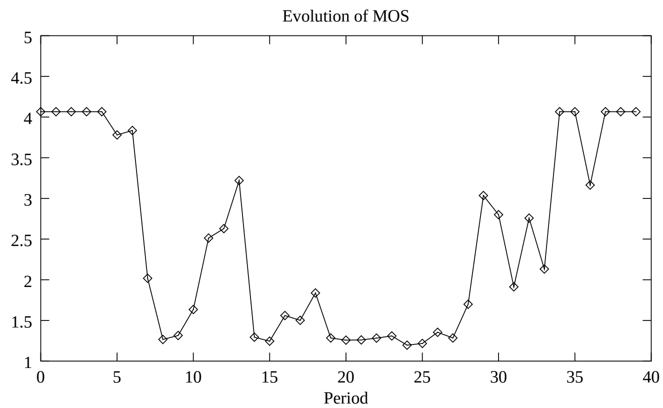

Figure 5: MOS scores for a VoIP flow (sampled every 5 seconds for real–time traffic) for the same 200s period.

These kind of results are not acceptable for normal use, since about half of the conversation would be incomprehensible. Even when some form of error correction, such as the forward error correction (FEC) scheme described in [12] is used, the MOS scores remain too low, as can be seen in Figure 6. This is due both to the high loss rate, and to the high values of the mean loss burst size, which sharply decreases the efficiency of FEC.

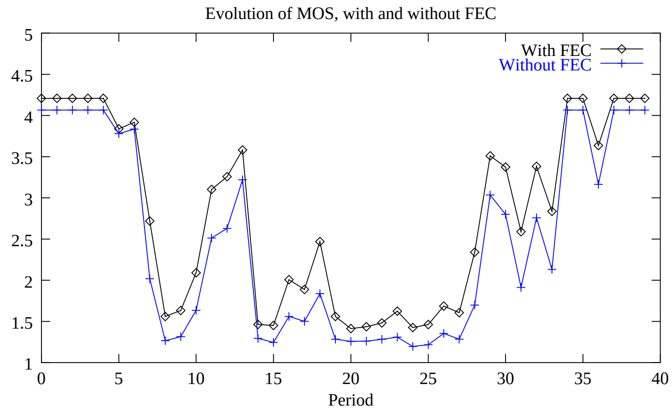

Figure 6: MOS scores for a VoIP flow, with and without FEC (sampled every 5 seconds for real–time traffic) for the same 200s period.

We performed several simulation runs, and all of them presented similar characteristics. While the fraction of time spent under the acceptable score varied slightly, every run presented at least an important period during which the quality was unacceptable. Figure 7 shows another representative MOS trace in this context.

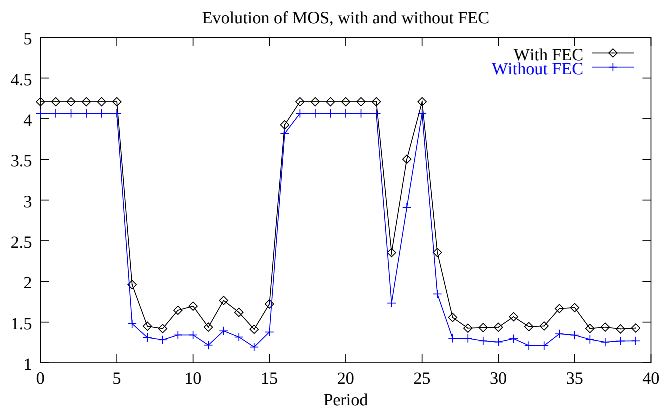

Figure 7: MOS scores for a VoIP flow, with and without FEC (sampled every 5 seconds for real–time traffic) for another 200s period.

4 Conclusions

We have studied the perceived quality of VoIP streams in a likely wireless home network scenario, where several devices are connected to the Internet and among themselves via a *wi-fi* network. We have proposed a stochastic model for the WLAN, and run several simulations in order to determine its performance.

It is easy to see that the variations in bandwidth that occur in *wi–fi* networks result in high loss rates. We used PSQA to determine how this would affect the quality of the voice streams as perceived by the end–user. According to our results, the quality is unacceptable, since during considerable periods of time the stream is incomprehensible.

We are currently studying mechanisms to improve the perceived quality, be it via adjustment of coding parameters (such as the FEC offset, for example), or at the network level, by introducing some form of service differentiation in the WLAN, be it at the IP level, or at the MAC level via the IEEE 802.11e [13] QoS mechanisms.

References

[1] S. Mohamed, G. Rubino, and M. Varela. A Method for Quantitative Evaluation of Audio Quality Over Packet Networks and its Comparison with Existing Techniques. In *Proceedings of the Measurement of Speech and Audio Quality in Networks workshop (MESAQIN'04)*, Prague, Czech Republic, June 2004.

[2] G. Rubino and M.n Varela. A New Approach for the Prediction of End-to-end Performance of Multimedia Streams . In *Proceedings of the First International Conference on Quantitative Evaluation of Systems (QEST'04)*, Enschede, The Netherlands, September 2004.

[3] M. Keppes and S. Garg. Can I Add a VoIP Call? In *ICC'03*, 2003.

[4] P. Mlhethaler. *WiFi 802.11 et les rseaux sans fil*. Eyrolles, 2002.

[5] M. Heusse, F. Rousseau, and A. Duda. Performance Anomaly of 802.11b. In *Infocom'03*, March 2003.

[6] R. Rollet and C. Mangin. IEEE 802.11a, 802.11e and Hiperlan/2 Goodput Performance Comparison in Real Radio Conditions. In *GlobeCom'03*, 2003.

[7] G. Rubino and M. Varela. Evaluating the Utility of Media–dependent FEC in VoIP Flows. In *QofIS'04*, Barcelona, Spain, September 2004.

[8] J. Incera and G. Rubino. Bit-level and Packet-level, or Pollaczec-Khintine formulae Revisited. In *QEST'04*, 2004.

[9] S. Mohamed, G. Rubino, and M. Varela. Performance Evaluation of Real-time Speech Through a Packet Network: a Random Neural Networks-based Approach. *Performance Evaluation*, 57(2):141–162, 2004.

[10] ITU-R Recommendation BT.500-10. Methodology for the Subjective Assessment of the quality of Television Pictures. In *International Telecommunication Union*, March 2000.

[11] ITU-T Recommendation P.800. Methods for Subjective Determination of Transmission Quality.

[12] J-C. Bolot and A. Vega Garcia. The case for FEC-based error control for packet audio in the Internet. In *ACM Multimedia Systems*, 1996.

[13] IEEE 802.11 Working Group. Draft amendment to standard for information technology telecommunications and information exchange between systemslan/man specific requirements part 11 wireless medium access control (mac) and physical layer (phy) specifications: Amendment 7: Medium access control (mac) quality of service (qos) enhancements, 2004.