A Method for Classifying VoIP System Impairments

Doris Bao[1,2], Luca De Vito[1,2], Laura Tomaciello[1,2]
[1] *Dept. of Engineering, University of Sannio, Corso Garibaldi, 107, 82100 Benevento, Italy*
Ph.: +39 0824 305817, Fax: +39 0824 305840,
E-mail: {doris.bao, devito, laura.tomaciello}@unisannio.it, http://lesim1.ing.unisannio.it
[2] *Telsey Telecommunications SpA, Benevento Research Laboratory,*
via dei Sanniti, 1, 82018 San Giorgio del Sannio (BN), Italy.
Ph.: +39 0824 330145, Fax: +39 0824 330145, http://www.telsey.com

Abstract - The voice quality in Internet Protocol (IP) telephony is significantly influenced by various factors, including delay, delay variation, bandwidth congestion and the distribution of packet loss. Hence, the performance of IP telephony systems is highly dependent on understanding the contribution of these factors to the voice quality. This paper deals about a method for classifying Voice over IP (VoIP) network impairments analyzing the signal power during a phone call. In particular, three main impairments have been considered: packet loss, jitter and bandwidth congestion. From the voice signal some figures of merit have been extracted and used to identify the impairments affecting the communication.

I. Introduction

The recent years have seen the growth of Internet Protocol (IP) based networks (e.g. Internet) at a thriving pace. The rapid proliferation and ubiquitous nature of the Internet, for example, has now given rise to strong interest in using IP based networks for carrying non conventional information like the voice, multimedia, etc. However, the use of the Internet, as a transport network for speech signals, is currently in its infancy [1].

VoIP (Voice over IP) is a time sensitive application, which transmits the human voice by packetizing the voice signal into a sequence of IP packets and reassembles them at the destination [2]. The voice quality in IP telephony is significantly influenced by various factors, including delay, delay variation, rate and distribution of packet loss. Hence, the performance of IP telephony systems is highly dependent on understanding the contribution of these factors to the voice quality [3].

Different methods known in literature are used to measure voice quality. Voice signal processing algorithms such as P.861 Perceptual Speech Quality Measure (PSQM) [4] and P.862 Perceptual Evaluation of Speech Quality (PESQ) [5] have been recommended by the ITU-T. These algorithms have been implemented in commercially available products that are used in the field to measure the end-to-end speech quality of VoIP calls. Such measurements are carried out by transmitting a packetized reference speech signal and then applying PSQM or PESQ to the received signal and the original reference to obtain an objective quality score. Alternative solutions, as shown in [6], operate directly on VoIP packet streams. Moreover, other methods either perform a voice quality analysis without suggesting the causes of speech degradation [7], or monitor the packet transmission but do not value the user perception [8].

This paper proposes a method, which allows retrieving the causes of the voice degradation by observing the voice signal. In particular, the method evaluates the VoIP network performance analyzing the power loss distribution during a telephone conversation. After a short description about a VoIP architecture, reported in Section II, the remainder of this paper is organized as follows: in Section III the impact of the network characteristics on the voice quality is presented; in Section IV the impairment classification method is described; Section V contains a detailed description of the test bench used to verify the method performance and some experimental results. Finally, the conclusions are presented.

II. VoIP architecture

An overview of the architecture of a VoIP system is described in Fig.1. In this figure, the four main components of a VoIP architecture based on H.323 protocol [9] are clearly shown: the terminal, the gatekeeper, a gateway and a Multipoint Control Unit (MCU). The terminal comprises the user terminal equipment and supports voice, video and data communication. The gatekeepers are responsible for translating between telephone numbers and IP addresses. They also manage the bandwidth and provide a mechanism for terminal registration and authentication. The gateway is the interface between heterogeneous networks such as the Local Area Network (LAN) and the Integrated Services Digital Network (ISDN) net. The MCU is needed to support conferencing [10].

III. Voice quality degradation factors

In a VoIP system, the voice transmission over the network is subjected to a degradation of the voice quality at the receiver. The interactivity between the communicating parties can be affected by the delays incurred in the network [11]. Indeed, a large delay may lead to "collisions" whereby participants talk at the same time. To avoid such collisions, the participants can talk in turns, and thus, take longer to complete the conversation. To achieve a good level of interactivity, the end-to-end delay (from "mouth to ear") should be maintained below a certain maximum delay. It is generally agreed that 150ms is the maximum desired one-way delay for an acceptable two-way communication [12]. Delays above 400ms result in a poor quality. When the delay becomes larger, there is a loss of synchronization between the speakers and the conversation begins to sound like two parties talking on a Citizens Band radio [13]. Longer delays become noticeable, and longer the end-to-end delay is, lower the degree of interactivity is.

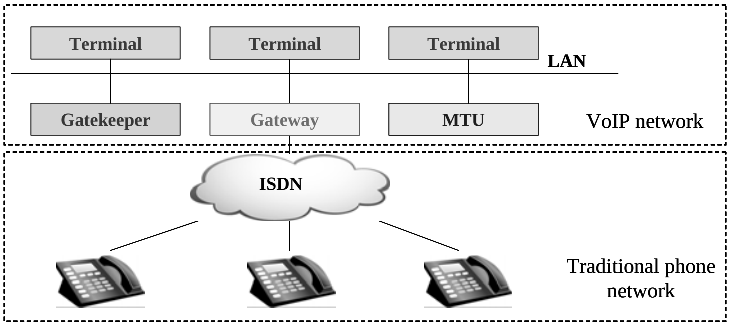

Fig.1. Architecture of a VoIP system.

Moreover, the network is subjected to the packet loss occurring in network elements, such as routers, whenever the buffer of the network elements is full or caused by many other factors including the jitter and the network congestions.

Since no network can guarantee a perfectly steady stream of packets under real-world conditions, VoIP phones use jitter buffers to smooth out the kinks. A jitter buffer is simply a First-In First-Out memory cache that collects the packets as they arrive, forwarding them to the codec evenly spaced and in proper sequence for accurate playback [14].

Specifying the length of the jitter buffer, and consequently the amount of delay, is fundamentally a trade-off:

– a large buffer allows more packets to be included in reproducing the speech but it adds a long delay;
– a short buffer increases the number of packets that are excluded from the generated speech. In fact, if the delay variation exceeds the size of the jitter buffer, there will be buffer overruns at the receiving end [15].

The overwhelming cause of loss can be, moreover, due to a congestion at the routers. It is, therefore, not surprising that there is a correlation between the bandwidth available and the amount of loss experienced [16-18].

As it is known, the effect of these network characteristics can be led back to a power loss of the voice signal and brings to a degradation of the received speech quality [19].

IV. Impairment classification method

The proposed method consists of the individuation and classification of the network impairments analyzing the voice signal. In this paper the most common network impairment are considered and, they are grouped on the basis of their effect on the VoIP signal: (i) random packet loss, (ii) jitter, and (iii) bandwidth congestion. The first determines a signal degradation due to temporary network conditions, when the packet is not received or it is discarded by the VoIP device. The second introduces an high variability on the packet arrival time. Finally, bandwidth congestion of one of the network devices, determines periodic packet losses, due to the continuous filling and emptying of the network device tails.

The idea proposed in this paper (Fig.2) is to send a voice test signal, known to the transmitter and the receiver, through the network and to analyse the power loss distribution of the received signal. In particular, an analyzer unit sends to the generator a request of a known test sequence. A generator unit sends the requested test sequence through the network. Then, the analyzer compares the

Fig.2. Impairment classification method.

digital voice signal on own side with that of the original test sequence.
A cross correlation between the generated signal and the acquired one is made to individuate the start instant of the conversation. From the comparison between the power levels of the two signals the I_l intervals of signal samples related to the power losses of the received voice signal are identified.
After that, to each sample *n* acquired, a value *v(n)* is associated, as shown in formula (1). In particular *v(n)* is 0 if the sample n belongs at least to an I_l interval and 1 otherwise:

$$v(n) = \begin{cases} 0 & \text{if} \quad n \in \bigcup_l I_l \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

In the network congestion case, the sequence of *v(n)* values has a particular periodicity because the number of the lost packets and the delivered ones repeats always in the same way. This is due to the continuous filling and emptying of the network device tails that determined the congestion.
The *v(n)* sequence is also used to calculate the duration of each I_l interval and to generate an histogram, h_i, where $i=1,\dots,N$, N is the number of classes and the amplitude of each class is L_C. If *P(00)* is the joint probability that one interval I_l has $2L_C$ samples, it can be estimated, as shown in equation (2):

$$P(00) = \sum_{i=1}^{N} h_i \cdot (i-1) \tag{2}$$

The formula (2) expresses the dependence between the power losses in consecutive temporal intervals and, therefore, it can be used to classify the received audio degradation due to random packets loss or to a high jitter.
The method has been realized using the scenario shown in Fig.3. It is composed of two units: a generator and an analyzer, both equipped with a Data Acquisition (DAQ) board. The analyzer unit requests to the generator a particular known test sequence. The generator accepts the request and sends the test sequence through the input DAQ to the network.
In order to implement the generator and analyzer units, LabVIEW 7.1 has been used. The generator unit, as shown in Fig.3, records the test sequences through a DAQ and saves them in a database. When the analyzer unit requests a particular signal, the generator finds and sends the requested test sequence. The analyzer unit, then, acquires and computes the received signal power loss.
A cross correlation between the generated signal and the acquired one is made to individuate the start instant of the conversation. Once detected such instant, the input and output instantaneous signal power are calculated and, through the comparison between the power values, the sequence *v(n)* and the histogram h_i are realized. Remembering that the speech signal alternates talk spurts and silence periods, the comparison permits to discriminate between the natural voice silence periods and silences due to packets loss: if the generated test sequence level power is high and the output test sequence is low, a power loss sample is counted; else it is a natural voice silence. Fig. 4 shows, in the first column, the power ratio between the received and the transmitted signal, in the second column, the FFT of the *v(n)* sequence and in the third column the histogram h_i, for three considered cases: 10% power loss, 150 ms jitter and 32kbit/s congestion, respectively.
To test all functionalities of the implementation, several experimental tests have been performed, introducing known impairments on the network, by means of a Linux PC, configured as a router, on which the network emulator NetEm has been performed. NetEm is an extension of the Linux “traffic control” tool. It allows defining packet scheduling policies that introduce impairments on the network, such as: packets loss, delay and jitter. The congestion has been emulated using the “traffic control” tool by limiting the router bandwidth using a “Token Bucket Filter” scheduling policy.
The network impairment classification has been performed by calculating the maximum value of the Fast Fourier Transform (FFT) of the sequence *v(n)* and the joint probability *P(00)*. The obtained values have been

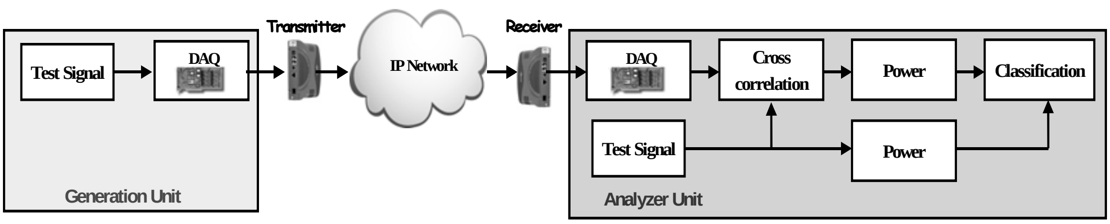

Fig.3. Generator and Analyzer Units block scheme.

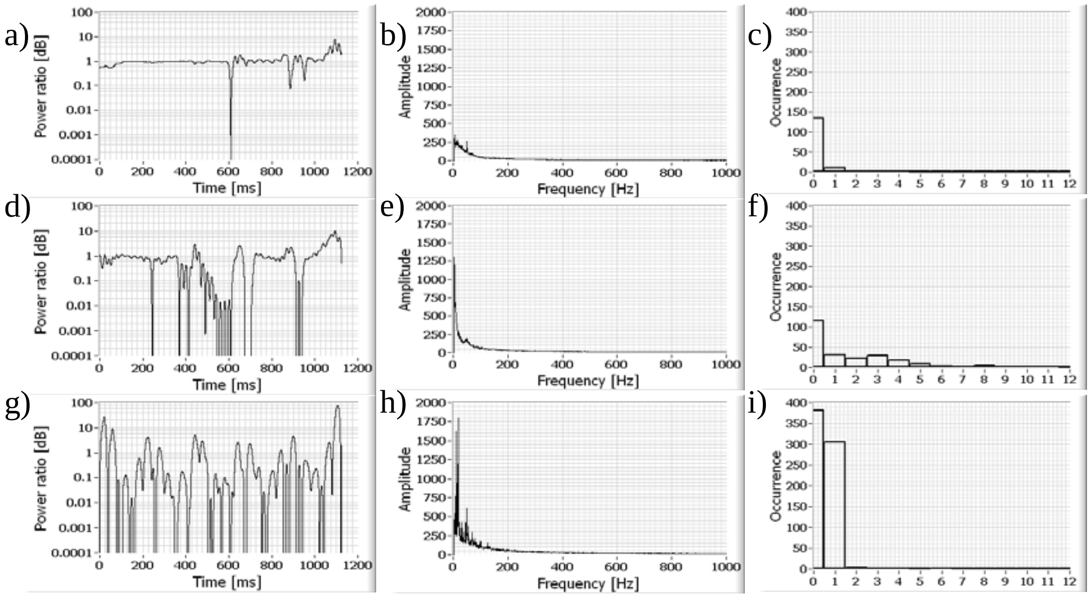

Fig. 4. Classification process: power ratio between received and transmitted signal, FFT of *v(n)* and histogram h_i for power loss 10% (a,b,c), jitter 150 ms (d,e,f) and congestion 32kbit/s (g,h,i).

compared with two thresholds, determined from an experimental phase, which allow the system to mark congestion, jitter or packet loss impairments.

A first threshold (Fig.5) has been calculated as the average among the least-square regression obtained by introducing in the network different bandwidth limit values (16kbit/s, 24kbit/s, 32kbit/s, 48kbit/s, 64kbit/s) and the linear regression on the maximum values of jitter and packet loss.

The second threshold has been calculated to differentiate jitter and packet loss. The *P(00)* values are lower than 0.1 in the case of random packet loss and higher than 0.1 in the case of jitter. In fact, when the impairment is a packet loss, the probability to have a power loss of the voice signal in two consecutive interval is low. Thus the threshold, *P(00)*= 0.1, is used to classified packet loss or jitter impairment.

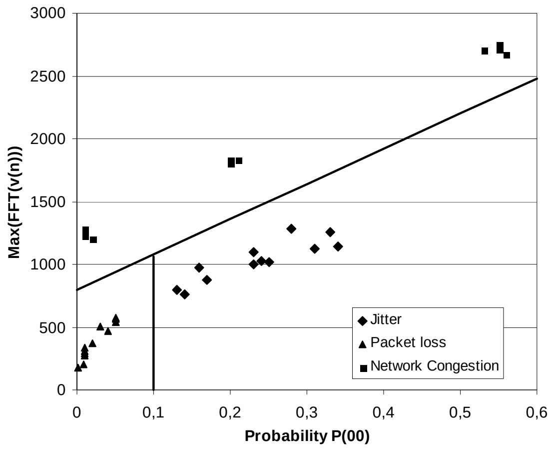

Fig.5.Threshold determination for the impairment classification.

V. Experimental results

In order to emulate a VoIP call, at Laboratory of Signal Processing and Measurement Information (LESIM) of the University of Sannio [20] a test plan has been set-up (Fig.6). As specified by the H.323 protocol, a simple call is established as it follows. If a user A wishes to talk to another user B, firstly A sends an admission request to its gatekeeper. The gatekeeper finds out, in the network, the IP

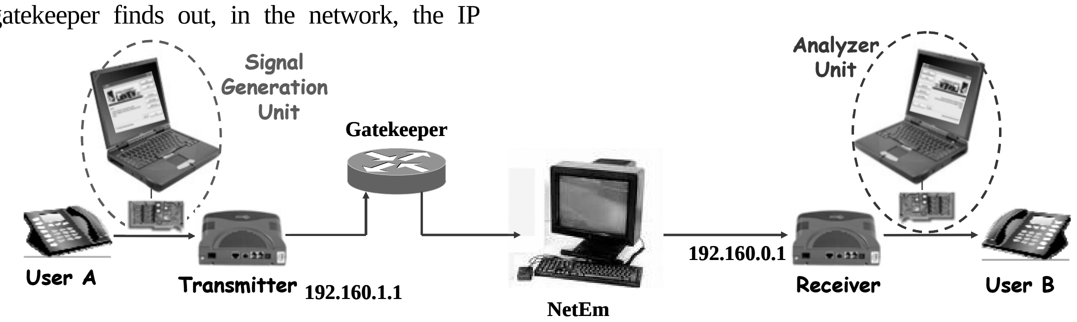

Fig.6. Test plan.

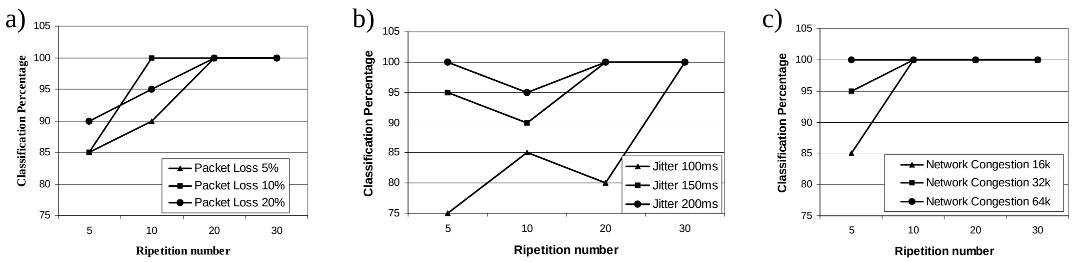

Fig.7. Percentage of correct classification: a) packet loss, b) jitter, and c) bandwidth congestion.

addresses at which B can be reached and informs A. After that, A establishes a TCP connection to the IP address of B, and the call begins. To emulate Internet, between two endpoints, constituted by two VoIP gateway and two telephones, a Linux PC running NetEm was located [21]. In such way, a controlled test environment has been created, emulating conditions found in a larger network, in which packets, in this case voice signals, may experience greater Quality of Service (QoS) degradations than in a simple test environment.

During a call, endpoints send call-signalling messages to the gatekeeper which routes the call to the destination endpoints through the network. The real or the emulated network add impairments on speech signal.

The experiments were realized emulating the different properties of a Wide Area Network (WAN): packet loss, variable delay and bandwidth congestion. The generated signal is an audio signal of 1.13s repeated 5, 10, 20 and 30 times. For each impairment value, 20 measurements have been performed.

In Fig.7 the percentage of the correct classification are reported. In each graph of Fig.7, the best classification is obtained for 30 iterations where, for each implemented impairment, the method classifies correctly the network impairment (Table 1). Instead, in the case of 5, 10 and 20 iterations the percentage of correct classification is equal to 75%. As shown in Fig.8, the FFT and P(00) measurements, obtained in the case of 30 iterations, are always in zones delimited by the implemented thresholds. Then, the maximum percentage classification value is obtained when the method acquires more samples and the acquisition time is longer as shown in Table 2.

Table 1. Percentage of correct impairment classification on 30 iterations.

	Packet loss			**Jitter**			**Bandwidth Congestion**		
	5%	**10%**	**20%**	**100 ms**	**150 ms**	**200 ms**	**16 kbps**	**32 kbps**	**64 kbps**
Packet loss	100	100	100	-	-	-	-	-	-
Jitter	-	-	-	100	100	100	-	-	-
Bandwidth Congestion	-	-	-	-	-	-	100	100	100

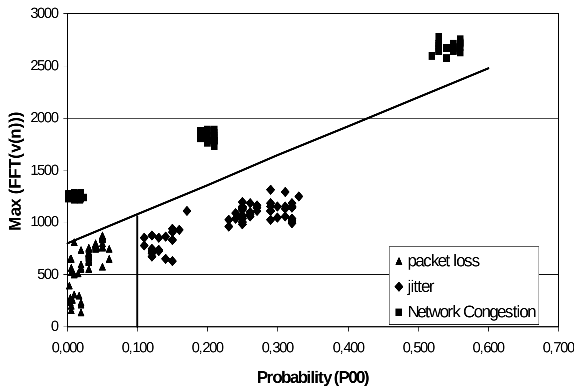

Fig.8. Impairment classification after 30 iterations.

Table 2. Classification time.

	Number of iterations			
	5	**10**	**20**	**30**
Packet loss	15s	37s	77s	113s
Jitter	12s	30s	64s	82s
Bandwidth Congestion	14.2s	31s	64s	98s

VI. Conclusions

This paper presents a method to classify the VoIP network impairment analysing the voice signal. In particular, analyzing the power loss distribution during a telephone conversation, the network characteristics degrading the voice signal have been individuated and classified. Three different type of impairments have been taken in consideration: packet loss, jitter, and bandwidth congestion.
The method showed good results in the experimental phase conducted using an emulated network. Further work are devoted to better characterize the method, even in a real world scenario.

Acknowledgements

The authors would like to thank Prof. P. Daponte for his helpful suggestions during all the phases of the research.

References

[1] M.K.Ranganathan, L.Kilmartin, "Performance analysis of secure session initiation protocol based VoIP networks", *Computer Communications*, vol. 26, No. 6, pp.552-565, Apr. 2003.

[2] P.Pragtong, K.Ahmed, T.Erke, "Analysis of speech data rate in voice over IP conversation", *IEEE TENCON*, Thailand, vol.1, pp.143–146, Nov. 2004.

[3] L.Roychoudhuri, E.Al-Shaera, G.B.Brewstera, "On the impact of loss and delay variation on Internet packet audio transmission", *Computer Communications*, vol. 29, No.10, pp.1578-1589, June 2006.

[4] ITU-T Recommendation P.861, "Objective quality measurement of telephone-band (300-3400 Hz) speech codecs", 1998.

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

[6] A.E. Conway, "A passive method for monitoring voice-over-IP call quality with ITU-T objective speech quality measurement methods", *IEEE International Conference on Communications (ICC)*, vol.4, pp.2583-2586, 2002.

[7] S.Lingfen, E.C.Ifeachor, "Voice quality prediction models and their application in VoIP networks", *IEEE Transactions on Multimedia*, vol.8, Issue 4, Aug. 2006, pp. 809-820.

[8] J.M.Matta, A.Takeshita, "End-to-end voice over IP quality of service estimation through router queuing delay monitoring", *Global Telecommunications Conference,* 2002. GLOBECOM '02. IEEE, vol.3, 17-21 Nov. 2002, pp.2458-2462.

[9] G.A.Thom, "H.323: the multimedia communications standard for local area networks", *IEEE Communications Magazine*, vol.34, pp.52-56, Dec. 1996.

[10] T. Uhl, "Quality of Service in VoIP communication", *AEÜ – International Journal of Electronics and Communications*, vol.58, No.3, pp.178-182, May 2004.

[11] P.Markopoulou, F.A.Tobagi, M.J.Karam, "Assessing the quality of voice communications over Internet backbones", *IEEE/ACM Trans. on Networking*, vol.11, No.5, pp.747-760, Oct. 2003.

[12] L.De Vito, S.Rapuano, L.Tomaciello, "One-Way Delay measurement: state of art", *Proc. of IEEE IMTC/2006*, Sorrento, Italy, pp.218-223, Apr. 2006.

[13] Goode, "Voice over Internet Protocol (VoIP)", *Proc. of IEEE*, vol.90, No.9, pp.1495-1517, Sep. 2002.

[14] Telchemy, Monitoring and Managing Voice over Internet Protocol (VoIP), www.networkinstruments.com, 2006.

[15] White Paper by NetPredict, "Assess the ability of your network to handle VoIP before you commit", Apr. 2002.

[16] C.Perkins, O.Hodson, V.Hardman, "A survey of packet loss recovery techniques for streaming audio", *IEEE Network*, vol.12, pp.40-48, Sep./Oct. 1998.

[17] S.B.Moon, J.Kurose, D.Towsley, "Correlation of packet delay and loss in the Internet", *Technical Report*, University of Massachusetts, 1998.

[18] K.Ishibashi, M.Aida, S.Kuribayashi, "Estimating packet loss-rate by using delay information and combinated with change-of-measure framework", *IEEE GLOBECOM*, pp. 3878-3882, 2003.

[19] J.Kim, S.Ho Han, H.W.Lee, W.Ryu, M.Hahn, "QoS-Factor transmission control mechanism for Voice over IP network based on RTCP-XR Scheme", *IEEE 10th ISCE*, pp.1-5, Jun. 2006.

[20] Laboratory of Signal Processing and Measurement Information - Department of Engineering, University of Sannio, Italy, http://lesim1.ing.unisannio.it.

[21] S.Hemminger, "Network emulation with NetEm", *Proc of Linux Conference*, Australia, Apr. 2005.