1. Introduction

The first step for the integration of multimedia traffic within the Internet is the provision of voice services, both traditional POTS (Plain Old Telephony Services) and innovative, like multi-party audio conferences or high quality audio. Voice services in Internet are broadly indicated as VoIP (Voice over Internet Protocol), including all possible services that can be built on vocal interaction, or simply on any elementary set of provisions for the transmission of audio and voice signals (e.g., CD-quality audio retrieval from archives). VoIP services can be offered on intranets (such as a corporate network), extranets (e.g., the network of a long distance PSTN – Public Switched Telephone Network – provider), or the global Internet. VoIP applications are deemed to play an important role in the future Internet architecture, while their QoS (Quality of Service) requirements do have a major impact on the design of both the user-to-network interface and the core network architecture. In other words, VoIP requirements are one of the inputs needed for the correct design of the Internet IntServ/DiffServ architecture, that is emerging as the standard reference frameworks in Internet research.

Research on VoIP comprises several aspects, ranging from voice encoding, to the definition of stochastic models that can be used to design and dimension the network properly (not to mention to allow efficient call admission control procedures), to the protocol used to transfer the voice or audio signals over IP networks. All the three aspects of VoIP source modeling outlined above have been tackled by the source modeling workgroup, subdividing the tasks between the three participating Universities.

2. VoIP Multi-Rate Speech Encoding

The first research issues faced by the research group are related to speech processing, in particular concentrating on aspects known to have a considerable impact on the QoS rendered by VoIP systems. The specific aim is the study and experimentation of variable bit rate speech coding techniques. In the first year, the activity focused on two issues:

- Improving phonetic speech classification algorithms;
 defining multi-rate coding systems.

As the improving of classification is concerned, the starting point is the ITU-T G.729 8 kbit/s speech coding standard. As a first step, a phonetic classifier with layered architecture was introduced, identifying 8 phonetic classes: 3 for voice activity and 5 for background noise. The followed approach is based on traditional pattern recognition; however, a new classification method was used in the matching phase which is based on fuzzy rules obtained by supervised training [1]. In order to make the phonetic classifier robust to environmental noise, a tool was developed in Matlab to order and select the best acoustic parameters to be extracted from the speech waveform. The subsequent investigation focused on new methods to enhance the performance of the Voice Activity Detector (VAD), which is of critical importance in Variable Bit Rate (VBR) speech coding. Two new approaches were studied, the first one using a noise suppressor before the VAD and the second one using a double microphone system; this latter is based on the difference in power levels and an estimate of the delay between the two inputs [2].

Appropriate coding models were studied for the 8 phonetic classes the codec provides for. A new algorithm called Backward Cross-Correlation (BCC) was developed to code fully voiced sounds. The algorithm exploits the correlation between the prediction residual in adjacent frames in fully-voiced sounds, reducing the bit rate to just 2.5 kbit/s in this coding mode. A subsequent study was carried out to characterise the prediction residual generated by several kinds of background noise in order to identify an appropriate LPC (Linear Predictive Coding) residual model. This study showed which types of background noise have a flat-spectrum or variable residual. In order to improve the coding gain, an ad-hoc codebook was developed for the coding of background noise with a variable-spectrum residual, with the aim of finding a trade-off between bit rate and voice quality in the reconstruction of the coded signal [1].

More recently the group focused on the characterisation, classification and suppression of background noise in order to study comfort noise coding models which will reduce the effect of discontinuity between periods of activity (ON) and inactivity (OFF) within a conversation. This method allows for the adaptation of the relative algorithms dynamically, according to the type of background noise [3, 4].

Finally, a series of comparisons were made between the developed VBR coding algorithm and the 8 kbit/s G.729 standard (in ON-OFF mode with VAD) with varying types of background noise and signal-to-noise ratios. The results obtained in terms of CMOS (Comparison Mean Opinion Scores) indicate that the performance of the two coding schemes in terms of QoS is equivalent, with the advantage of a reduction in bandwidth ranging between 10 and 20% depending on the kind of background noise.

In parallel, the research group worked on a multi-rate version of the encoder, which should be less sensitive to problems linked to the transmission of speech on IP networks. More specifically, scalability has been introduced by inserting two further modes – 6.4 and 11.8 kbit/s – which represent extensions to the 8 kbit/s G.729 codec recently standardised by ITU-T. In this way it is possible, by means of a rate control mechanism, to adapt the peak bit rate according to the network load or the percentage of lost packets.

The research activity will now focus on subjective evaluation of quality in terms of CMOS. An IP network simulator will be used to generate controllable network situations, trying to devise objective evaluation methods by identifying parameters that are correlated with subjective measurements. The performance of the new coding algorithms and that of
traditional algorithms will be compared in various network resource and background noise conditions.

3. Stochastic Modeling and Characterization of VoIP

The second key point of the research is dedicated to stochastic source modeling. As widely recognized, a key challenge is represented by developing new architecture paradigms for Internet, whose aim is the satisfaction of QoS requirements of innovative IP-based services: VoIP the first among them. The relevance of this issue is related to the metamorphosis of Internet into a commercial infrastructure providing different services to users with different service requirements. In this scenario, the availability of an architecture paradigm providing different QoS classes is just as important as increasing the bandwidth availability. Among several proposals, the DiffServ approach seems the most promising for the implementation of scalable service differentiation in the Internet. Scalability is achieved by assuring the service only over traffic aggregates (whose profile is more predictable), while single microflows are disregarded in the network core. At the same time, the traffic injected at the edge of the network is conditioned to meet the required profile. DiffServ aims at the provisioning of QoS through a small, well-defined ensemble of building blocks enabling a large variety of services [6,7]. These building blocks include a set of per-hop forwarding behaviors plus packet classification and traffic conditioning functions such as metering, marking, shaping and policing. In this framework, the selection of a proper traffic descriptor (which permits to specify the Traffic Conditioning Agreement) is fundamental both to achieve efficient resource allocation among the users and to correctly design scheduling algorithms providing differentiated QoS.

In the first year of the project our attention has been mainly focused on metering algorithms. Most researchers agree in considering the token bucket algorithm as an efficient meter; while the traffic profile can be described by means of parameters derived from LBAP (Linear Bounded Arrival Processes) traffic characterization [8]. A LBAP modeled source transmits, in any time interval of length \( t \), a number of bits that is \( \leq \lambda t + b \), where \( \lambda \) and \( b \) are the two parameters characterizing the traffic. The aim of the analysis carried out during the project is the evaluation of the operating parameters of the LBAP description starting from the traffic stochastic model. The proposed analysis, based on an equivalent queuing system, can be applied using different traffic models, although practically only few cases allows for an easy solution of the problem. The analysis was focused on a case study representing a possible VoIP scenario. The traffic to be characterized is an aggregation of independent fluidic On-Off sources, where the permanence times in each state is exponentially distributed. The On-Off assumption is due to the typical behavior of a voice source with VA: it is active or inactive depending on whether the talker is speaking or silent. Assuming that no compression is applied to voice signal, during active periods the source transmits at the constant bit rate of 64 Kbps (this corresponds to a standard PCM codec with silence suppression). Moreover, in-depth analyses of this traffic source available in the literature [9], have emphasised that the distribution of active and inactive periods lengths can be approximated by an exponential function.

Theory developed in [5,10] for fluidic On-Off sources provides an efficient computational procedure for \((\lambda, b)\) evaluation. The errors introduced by the fluidic approximation were
evaluated by means of discrete event simulations carried out using the OPNET network simulator.

4. Rate Adaptive Protocols for VoIP Applications

The last item is related to end-to-end protocols for voice support. The research focused on adaptive techniques allowing a source to vary its transmission rate in response to end-to-end congestion indicators, such as the percentage of lost packets, the packet delivery delay and its variance.

The research followed three main directions, methodologically complementing one another:

1. Definition of an analytical framework [11] to study the influence of several parameters such as delay and loss on system performance. A detailed Markovian description of the variable-bit-rate source was introduced, while an approximate description of the network-source interaction was estimated through a fixed-point procedure. This model has provided us with performance indices in a case study where the statistical characterization of the average source transmission rate was evaluated, along with estimates of average packet delay and loss rate;

2. Simulation study of adaptive algorithms. Simulation allowed us to test the efficiency of several rate control algorithms, and the role of threshold parameters governing the operation of the algorithms. The simulator was built from an existing, well-established simulator for packet-switched networks, called ‘ns’;

3. Experimental tests (see [12] for details) using a software tool for the coding and transmission of adaptive Voice over IP. The tool was developed for Windows and Linux platforms (although the latter still is in a beta testing phase).

The three methodologies have highlighted significant gains in terms of delay and percentage of lost packets when variable-rate and constant-rate techniques are compared, especially when there is no aprioristic notion of the available bandwidth, as it is often the case.

References


