Understanding VoIP from Backbone Measurements

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2007 proceedings.

Understanding VoIP from Backbone Measurements

Robert Birke, Marco Mellia, Michele Petracca
Politecnico di Torino – Dipartimento di Elettronica
name.surname@tlc.polito.it
Dario Rossi
ENST Télécom Paris - GET/ENST/INFRES/RMS
dario.rossi@enst.fr

***Abstract*— VoIP has widely been addressed as the technology that will change the Telecommunication model opening the path for convergence. Still today this revolution is far from being complete, since the majority of telephone calls are originated by circuit-oriented networks. In this paper for the first time to the best of our knowledge, we present a large dataset of measurements collected from the FastWeb backbone, which is one of the first worldwide Telecom operator to offer VoIP and high-speed data access to the end-user. Traffic characterization will focus on several layers, focusing on both end-user and ISP perspective. In particular, we highlight that, among loss, delay and jitter, only the first index may affect the VoIP call quality. Results show that the technology is mature to make the final step, allowing the integration of data and real-time services over the Internet.**

I. Introduction

The evolution of the Internet toward a universal communication network has been foreseen by both researchers and Telecom providers. Voice over IP (VoIP) has long been indicated as the technology that will trigger this revolution, definitively opening the path for convergence. VoIP technology is available since more than fifteen years, and standards are available since the mid of '90s considering both signaling [1], [2] and transport protocols [3], as well as voice Codecs [4]. Albeit networking technology has evolved, offering both users and Telecom providers high-speed access and backbone networks, still today the revolution is far from being complete. Indeed, while the Internet has definitively been accepted as the only data communication network, the large majority of voice traffic is originated from circuit-oriented networks through the old Public Switched Telephone Network (PSTN).

Traffic monitoring and characterization have always been seen as a key methodology to understand telecommunication technology and operation, and the complexity of the Internet has attracted many researchers to face traffic measurements since the pioneering times [5]. Data traffic has hogged the majority of this effort, while the attention toward VoIP traffic measurements only recently increased [6], [7], [8], [9], [10]. In particular, authors in [6] present a methodology to perform VoIP quality assessment over backbone networks. Active measurements are collected from a US backbone network, and then results are fed to a model to derive rating for equivalent VoIP calls. Similarly, in [7] authors consider the quality of VoIP calls offered by several VoIP applications running on PCs and VoIP phones connected to a simple LAN environment. Similar results are presented in [8], in which authors consider a H.323 compatible setup, and define a mapping between network layer measurements to VoIP quality. Testbed experiments are presented, in which VoIP artificial traffic is sent through a LAN and WAN environment. In [9] a passive methodology to monitor VoIP phone calls is described, but only simple measurements over an artificially loaded testbed are presented. Finally, authors in [10] study VoIP quality mapping considering the ITU-T E-Model [11], whose implementation is described in [12].

All previously mentioned works rely on traffic characterization and measurement obtained from active probes, in which controlled sources, either PCs, VoIP phones or traffic generators, are used to inject packets in LAN or simple WAN environments. To the best of our knowledge, no extended measurement study is available in the literature that is based on purely passive monitoring of VoIP traffic. In this paper, we present the first extended set of measurement results collected via passive monitoring of real-world VoIP traffic. Real traffic traces are collected from an ISP provider in Italy, called FastWeb [13], which is the main broadband telecommunication company in Italy, offering telecommunication services to more than 5 millions of families, with 1 million of subscribers (11% of market share). Thanks to its fully IP architecture, and the use of either Fiber to the Home (FTTH) or Digital Subscriber Line (xDSL) access, FastWeb has optimized the delivery of converged services, like data, VoIP, IP television, over a single broadband connection. No PSTN circuit is offered to end-users, so that native VoIP is adopted. Measurements cover both application/user layer indexes (such as phone call duration and perceived quality), and network layer parameters (such as loss probability, round trip time and jitter). To better understand the impact of the network access technology and network topology, results will be presented conditioning on homogeneous sets (e.g., discriminating users depending on their access technology or the cities they live in). Results show that the technology is mature to make the final convergence step, allowing the integration of data and real-time services over the Internet. In addition, among network layer performance indexes, the loss probability is found to be the principal cause of VoIP quality degradation in the FastWeb scenario. Moreover, besides contributing to the understanding of the

VoIP technology, all the algorithms described in this paper are made available to the research community via an open-source tool called Tstat - TCP STatistic and Analysis Tool [14].

The rest of the paper is organized as follows: after presenting the measurement methodology in Sec. II, the FastWeb architecture is detailed in Sec. III. Measurement results are reported in Sec. IV, distinguishing between user-centric and network-centric indexes in Sec. IV-A and Sec. IV-B respectively. Finally, Sec. V concludes the paper.

II. Measurement Methodology

In this Section, we define the measurement methodology adopted to perform the traffic characterization, focusing on multimedia streams in particular. We assume that a monitoring probe is used to sniff packet headers from traffic flowing on a link, so that the first bytes of the packet payload (up to part of the RTP/RTCP headers) are exposed to the analyzer. We also assume to observe a bidirectional stream of packets, so that both packets going to and coming from a node can be monitored.

All the developed algorithms have been implemented in Tstat [14]. Tstat is an IP networks monitoring and performance analysis tool developed by the Telecommunication Networks Group at the Politecnico di Torino. By passively observing traffic on a network link, Tstat computes a set of performance indexes at both the network (IP) and transport (TCP/UDP) layers. Originally focused on data traffic, Tstat has been enhanced to monitor multimedia streams, based on RTP/RTCP [3] protocols carried both over UDP or tunneled over TCP.

A. Identification of RTP/RTCP over UDP flows

VoIP and more generally multimedia applications usually rely on UDP at transport layer, which offers a connectionless, unreliable service. RTP/RTCP [3] are standard protocols used to support the additional features required to transport multimedia traffic over UDP, such as sequence identification, stream synchronization, etc. Both UDP and RTP/RTCP are characterized by the lack of connection signaling, therefore making it difficult to identify the connection setup, data transmission and tear-down phases. Indeed, several out-of-band signaling protocols can be used to setup a VoIP call, like the one supported by H.323 [1] or SIP [2] standards to name a few. In order to identify a multimedia call, two options are available. The first one is to identify and interpret the signaling connection, consequently snooping the associated data connection. The second one is to identify directly the data connection. While the first method would prove more reliable, it is much more complex (since it requires a complete knowledge of the signaling protocols adopted) and troublesome (since it requires that both voice and signalling flows are exposed to the probe point and the whole signaling packet payload is needed by the monitoring tool). We therefore propose a passive identification methodology of RTP/RTCP flows that relies on the observation and monitoring of data streams only. Notice that only packet headers are required to be exposed to the analyzer, so that

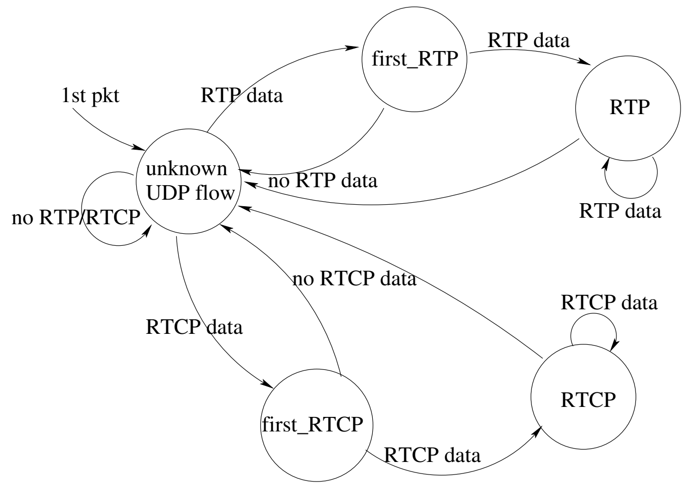

Fig. 1. Heuristic RTP over UDP identification

privacy concerns are limited. In particular, considering RTP, 12B of RTP headers are required, for a total of 40B of IP packet. Considering RTCP, up to 52B may be required to correctly decode a receiver report piggybacked to a sender report, giving a total of 80B of IP packet.

Being impossible to detect RTP/RTCP flows by, e.g., port numbers, we defined a heuristic algorithm based on the Finite State Machine (FSM) shown in Figure 1. Each UDP flow, identified by using the traditional tuple (source and destination IP, source and destination port, protocol type), is tracked. A new flow starts when a packet is first observed, while an inactivity timeout is adopted to define the flow end[1]. Notice that for each bidirectional flow of UDP packets, two half-duplex streams are identified, i.e., one for each direction between the two nodes.

When the first UDP packet is observed, the flow is labeled as *unknown*. For each new UDP packet, the algorithm double checks if the UDP payload may be identified as a RTP/RTCP message. According to [3], some fields of the RTP/RTCP headers must satisfy some assumptions. In particular:

- the version field must be set to 2;
- the payload type field must have an admissible value;
- the UDP port must be larger than 1024.

If all three conditions are satisfied, then the flow is marked as a possible RTP/RTCP flow (*first_RTP* and *first_RTCP* state respectively). The *first_RTP/first_RTCP* state is entered if an even/odd UDP port number is observed. The flow i) Synchronization SouRCe identifier (SSRC), ii) payload type, and iii) sequence number (in case of a RTP flow) are then initialized to the value observed in the first packet.

When the next UDP packet belonging to the same flow is observed, the FSM checks if, in the case of RTP:

- the version is equal to 2;
- the same SSRC of the first packet is present;
- the same payload type of the first packet is present;

[1]We conservatively set the timeout value to $200s$.

- the sequence number is the expected one;

or in the case of RTCP:

- the version is equal to 2;
- the same SSRC of the first packet is present;
- the packet type is an admissible one;

In case of correct matching, *RTP* or *RTCP* state are entered respectively, and the flow is labeled as RTP/RTCP and its analysis can start.

Then, for each observed UDP packet, the FSM verifies that the UDP payload still contains valid RTP/RTCP values. In case this is not true, the FSM discards the flow, moving back to the *unknown* state.

Every RTP/RTCP monitored flow will be correctly classified by the FSM. However, it is possible that an UDP flow which is neither a RTP nor a RTCP flow is positively identified by the FSM heuristic, causing a false-positive identification. Nevertheless, odds are such that misclassification is an extremely rare event: since the pattern used to perform the matching has a total equivalent length of 10bits considering the first packet, and 58bits considering the second packet, then the false-positive probability accounts to 2^{-68} in case random fields are assumed.

B. Measurement Indexes

Once the RTP/RTCP flows[2] are identified, a *phone call* is pinpointed by matching each RTP flow to its corresponding RTCP flow on the basis of the IP addresses, UDP port numbers and SSRC identifier, so that two RTP and two RTCP half duplex flows are grouped together. Due to routing asymmetry, and to possible source configuration, it is possible that some of the above mentioned flows are not present, e.g., no RTCP flow is present. Since we are interested in the quality of the phone call, we require all flows to be present – otherwise, we are forced to discard the incomplete sample. Also RTP/RTCP flows not belonging to phone calls are discarded based on the payload type.

Since both RTP and RTCP header information is available to the monitoring probe, several possibilities may be adopted to estimate the above mentioned quality parameters. For example, the loss probability may be inferred by monitoring the RTP *sequence number* field, or relying on the RTCP *cumulative number of packets lost* or *fraction lost* fields. After testing some possible techniques, we selected the most reliable one. In particular, we noticed that the information carried by RTCP reports is often very unreliable, possibly due to not fully standard implementation. On the contrary, RTP header information is much more reliable, and very rarely it is affected by possible implementation errors. Indeed, while RTCP carries only *control* information which may be used by the receiver, RTP packets must be correctly interpreted by the receiver, otherwise serious compatibility problems may arise. Therefore, whenever possible, we rely on direct measurements

[2]In the paper, we refer to a "flow" considering the monodirectional stream of packets flowing from a given source to a destination. Therefore, a phone call is build by four flows.

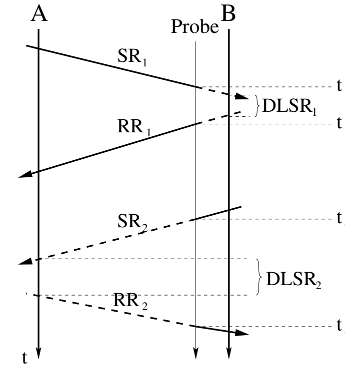

Fig. 2. Time diagram of RTT estimating

from RTP packets observation, rather than on client-based measurements reported in the RTCP packets. Whenever RTCP information had to be used, we discarded unreliable samples by discarding clearly "wrong" values (i.e., where fields have clearly other boundaries and interpretation than those defined in the standard, as we happened to observe negative times, total number of packet lost larger than the flow packets, average jitter values larger than 1s, etc.) In the following, a brief description of the algorithms used to perform the packet level measurements is given.

Considering the quality parameter, the following indexes have been monitored:

- *Call Duration* (τ)
- *Call Round Trip Time* (RTT)
- *Flow Packet Loss probability* (P_l)
- *Flow Jitter*
- *Flow equivalent Mean Opinion Score* (eMOS)

The call duration τ is defined as the time elapsed between the first and last RTP packet reception at the monitoring probe.

The RTT is defined as the time needed by a packet to be routed from the sender to the receiver and then back to the sender, which is impossible to be gauged by a single measurement point unless a feedback is immediately sent by the receiver back to the sender. While this is true considering reliable connection-oriented protocols (e.g., TCP, SCTP) in which acknowledgments are used, the estimation of the RTT is more complex in case of unreliable connectionless transport protocols. However, the RTP/RTCP protocol specification requires the receiver to send a report back to the source at periodic intervals: by analyzing RTCP Sender Reports (SRs) and Receiver Reports (RRs), it is then possible to estimate the RTT. As shown in Fig. 2, the monitoring probe is placed along the path between the two client nodes (named A and B in the Figure). The RTT can be split into two parts, considering the RTT from each source to the measurement point. Each semi-RTT, enclosed by two dashed lines in the Figure, is estimated from the observation of a SR-RR couple. Let t_a be the time when node A SR_1 is observed by the probe. When SR_1 has reached node B, RR_1 will be sent back to node A after an elaboration time, called *Delay since Last Sender Report* ($DLSR_1$). The $DLSR_1$ value will be written by node

B in the RR_1, and therefore will be available at the probe node at time t_b when the RR_1 will be observed. Similarly, the estimation of the semi-RTT among the probe and node A can be obtained by monitoring the time $t_c,\ t_d$ at which SR_2 and RR_2 are observed, and subtracting the elaboration time $DLSR_2$. Therefore we have

$$RTT = (t_b - t_a) - DLSR_1 + (t_d - t_c) - DLSR_2 \qquad (1)$$

Finally, the average RTT is evaluated by averaging among all RTT samples during call lifetime.

The P_l is the probability to loose a RTP packet during the call lifetime, evaluated as the fraction of lost packets over the total number of flow packets. To identify lost packets we implemented a sliding window mechanism recording the observed packet sequence number, relying thus on the monitoring of the RTP *sequence number* field, so that numbering gaps can be detected, as well as duplicate packets. A lost packet is identified if its sequence number has never been observed by the probe node, while subsequent packets have: the sliding window mechanism is introduced to limit memory usage at the probe node. In other words, a packet is considered as lost if it has never been observed (i.e., it is actually lost in the network) or if it is delayed so that it arrives at the probe outside the sliding window boundaries (i.e., a *late* packet is observed). We set the window size to 20 packets, or 400ms considering a packetization time of 20ms: notice that late packets are practically too late to be useful to the application.

The $jitter$ is the measurement of the Inter-Packet-Gap (IPG) variation. IPG is evaluated from the time-stamp of two consecutive packets belonging to the same RTP flow at the probe point. Then IPG samples are used to feed the jitter estimator, following the dictations reported in [3]. Notice that the jitter measurement is performed at the probe node, and is not equal to the jitter evaluated at the client nodes. However, being the probe very close to the destination in the measurements scenario considered in this paper, we can neglect the missing contribution to the jitter.

Finally, the *eMOS* is evaluated according to [11]. Since the evaluation of the MOS is rather complex, we report in Appendix the details and the adopted settings. The eMOS is a computational model standardized by ITU-T through recommendation G.107 [11] that predicts the subjective quality of packetized voice. The outcome can be further translated into an equivalent MOS, which provides a numerical indication of the perceived quality of received media after compression and/or transmission.

III. The FastWeb Network

FastWeb was born in October 1999 with a revolutionary idea of delivering only Internet access to end-users (both consumer, SOHO, and big customer) and providing telecommunication services over IP. In October 2000, the service was opened to consumers and business customers, offering Internet access, VoIP telephony and video on demand services. Since then, FastWeb has become the main broadband telecommunication company in Italy. Thanks to its fully IP architecture, and the use of either Fiber-To-The-Home (FTTH) or xDSL access technologies, FastWeb has optimized the delivery of converged services, like data, VoIP, IPTV, over a single broadband connection. In this Section we briefly introduce the FastWeb architecture, describing the access network, the backbone network and finally the VoIP architecture.

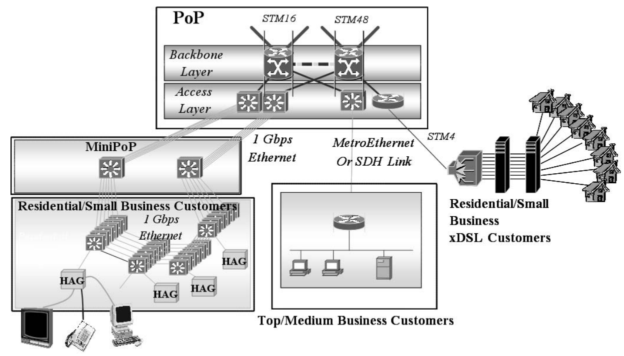

Fig. 3. The FastWeb infrastructure: FTTH and xDSL access, MiniPoP, PoP and backbone layers.

As presented in Fig. 3, a Metropolitan Area Network (MAN) Ethernet-based architecture is adopted in the last mile. Residential and small business customers are connected to a Home Access Gateway (HAG), which offers Ethernet ports to connect PCs and the VideoBox, as well as Plain Old Telephone Service (POTS) plugs to connect traditional phones. The HAG is essentially an Ethernet Switch, combined with a H.323 gateway to convert POTS analog input to VoIP transport. A 10Base-F port is used to connect the HAG to a L2 switch installed in the basement in case of FTTH access, while a modem port is used when xDSL access is offered. In the first case, L2 switches are interconnected by 1000Base-SX links forming bidirectional rings. Rings are terminated at the so called MiniPoP by means of two L2 switches, configured as a spanning tree root to recover from faults. A trunk of several 1000Base-SX links connects each MiniPoP switch to a L2 switch in the PoP, in which two routers are used to connect the backbone by means of Packet-Over-Sonet (POS) STM16 or STM48 links. In case of xDSL access, the HAG is connected to the traditional twisted pair phone link terminated directly to a DSLAM. Then either a STM4 or STM16 link is used to connect DSLAMs to the PoP by means of an additional router, as shown in the right part of Fig. 3; notice that no analog circuit is present even when using xDSL access. When FTTH access is adopted, customers are offered 10Mbps Half-Duplex Ethernet link, while when xDSL access is adopted, customers are offered 512Kbps or 1024Kbps upstream and 6Mbps or 20Mbps downstream link. Finally, medium/top business customers are offered both MetroEthernet or SDH access by means of a router connected directly at the PoP layer.

Cities covered by the MAN access infrastructure are interconnected by means of a high-speed backbone based on IP-over-DWDM technology. The largest cities in Italy are directly

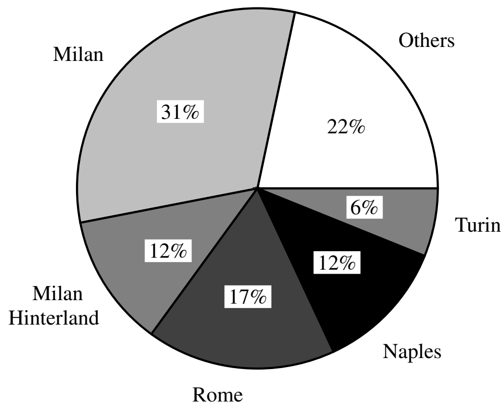

Fig. 4. Distribution of HAG location for the most active cities. gateway trace.

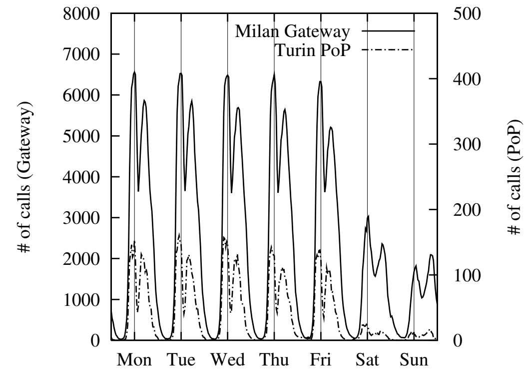

Fig. 5. Number of observed phone call versus time.

connected by more than 12.400km of optical fibers. In each city, one or more PoPs are present, while several MiniPoPs are installed so that each collects traffic from up to 10.000 users.

Considering the services offered to customers, FastWeb offers traditional data access, telephony, video on demand and multicast streaming of more than 100 digital TV channels. At the risk of being tedious, we recall that all services use IP at network layer. In particular, the VoIP architecture, which is the topic of the measurements in this paper, is based on both H.323 and SIP standards. The HAG converts traditional analog phone ports to VoIP and performs both signaling and voice transport tasks. Phone calls between FastWeb users are then routed end-to-end without any further conversion, while phone calls to traditional users are routed toward a gateway to be converted to the PSTN of Telecom Italia. One or more gateways are installed in the largest cities, so that long-distance calls are routed over the FastWeb IP network up to the closest gateway to the destination, and then converted to PSTN.

Considering the voice transport, a simple G.711a Codec without loss concealment is used, so that two 64kbps streams are required to carry the bidirectional phone call. Packetization time is set to 20ms, leading to 160B of voice samples per packet. RTP and RTCP over UDP are used to transport the voice streams. No per-class differentiation is performed by the network layer, so that VoIP, data and video streams are all multiplexed into a single aggregate stream.

IV. Measurement Results

In the following, we present results obtained monitoring traffic at both the MiniPoP level and the backbone gateway level. Two probe nodes based on high-end PCs running Linux have been installed in a PoP located in Turin, and in a Gateway node located in Milan. The first probe has been connected to one of the two PoP L2-switches, that was configured to replicate all traffic coming in and going out through the links connecting the PoP backbone router. Tstat was directly run on the probe so that live traffic measurements are taken, and results can be observed from the Web through Tstat Web interface [14]. An average load of 100Mbps full-duplex with peaks up to 500Mbps of traffic has been processed for more than three weeks. The mix of data and VoIP traffic was monitored.

Similarly, a second probe has been connected to a dedicated device that replicated all packets flowing on a full-duplex link connecting an array of gateways located in Milan. The gateways are used to route phone calls to and from the PSTN, so that only RTP/RTCP traffic is present on that link. The average load on the measurement link was 350Mbps, peaking to 600Mbps. Also in this second case, Tstat was running live for more than two weeks.

Given that the probe points are very close to one of the two call parties, flow measurements will be presented considering the monodirectional stream coming from the farthest party, i.e., flows from the HAG to the gateway in the Milan gateway probe, and flows to the local HAG in the Turin PoP probe.

In the following, we present results trying to highlight the differences among them. In particular, we will classify results[3] according to the i) xDSL or FTTH access offered by the HAG, ii) source cities location, discerning in particular Milan, Turin, Rome and Naples, iii) edge or gateway measurement points. The distinction will be made whenever a significant difference is appreciable. Considering the user network access, we have that FTTH technology is adopted by about 1/4 of HAG, while 3/4 of users are connected by means of xDSL access.

Results will be presented showing both the time evolution of the measurements (considering 5 minutes averaging intervals in the daily scale and 30 minutes on the weekly scale), and the corresponding Probability Density Function (PDF) or, equivalently, the Cumulative Distribution Function (CDF). PDF and CDF are obtained from the gateway probe considering a stationary interval of time. Indeed, results have been obtained processing a 4 hour-long packet level trace recorded from 10am to 2pm of Saturday the 15th of July 2006. Total trace length amount to about 240GB of packet headers, corresponding to more than 150.000 phone calls.

Fig. 4 reports the distribution of HAG location observed

[3]The classification has been made according to the source and destination IP addresses, which are assigned by FastWeb following a known addressing plan.

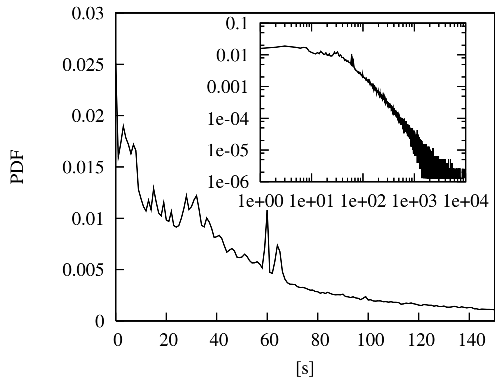

Fig. 6. PDF of phone call duration in linear scale and in log/log scale in the inset.

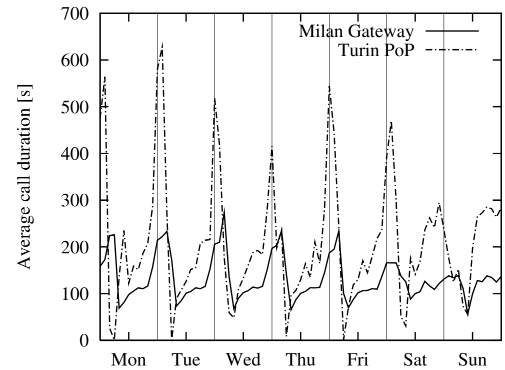

Fig. 7. Average phone call duration versus time.

from the gateway trace. Only the five largest destinations are reported, which correspond to Milan and its hinterland (aggregated to Milan in the following), followed by Rome, Naples and Turin. This does not exactly correspond to the equivalent ranking of the city according to their population, due to the bias of the geographical location of the monitoring probe.

To give the intuition of the number of phone calls tracked by the probes, Fig. 6 reports the evolution versus time of the number of calls monitored during the third week of July 2006. It is possible to notice the typical day/night periodicity, and also to identify week-days from weekend-days, during which the network load is smaller. The maximum values (more then 1300 simultaneous calls per minute considering the gateway probe) are observed between 10am to 2pm, while a sudden decrease is observed during lunch break and in the early afternoon. Considering the PoP trace, the qualitative results are very similar, but a scaling factor of about 30 is observed, so that two different axis are used.

Comparing these results to similar ones tracking the activity on data networks (see [16] for example), similar day/night trend is noticed, but no traffic reduction has ever been observed during the lunch break.

A. User-Centric Measurements

We start by showing the call length τ. In particular, Fig. 6 reports the PDF of τ using lin/lin scale in the large plot, and log/log scale in the inset. The overall average phone duration is 106s, with long-distance calls showing a smaller average (97s) compared to local calls (113s). It can be noticed that τ PDF follows a heavy tailed distribution, as underlined by the linear shape of curve in the log/log inset. A large part of the distribution mass is found in the $[0 : 10]$s range. This is possibly due to phone calls that were not answered being either the callee party busy or missing. There is also a noticeable peak at about 60s, but after investigating a possible motivation for this, we were not able to underline any particular cause.

Considering the average call duration versus time, Fig. 7 plots the results of one week of monitoring. It is possible to notice the difference between the average call duration during working hours [8am:6pm] and during the early evening and night. Indeed, the average phone call duration is much larger in the second part of the day, where shorter work-calls are substituted by longer friendly chatting. Another reason of the different average call duration observed at the PoP or at the gateway may be due to different pricing adopted by FastWeb: indeed, calls that are terminated at a PSTN phone are always charged, while calls between FastWeb users are free.

Finally, we then present the most interesting result, i.e., the quality of service faced by VoIP users, via the eMOS index. We recall that a value of the eMOS larger than 4 is considered "excellent quality" (no perceptible impairment), while a eMOS in $[3 : 4]$ range corresponds to "good quality" (perceptible but not annoying impairment). Finally eMOS equal to or larger than 3.6 is considered the same quality as traditional PSTN phone calls.

Fig. 8 reports the eMOS CDF considering different originating cities (top plot) and for different access network (bottom plot). It can be noticed that more than 60% of phone calls exibiths an excellent quality (eMOS> 4), while about 85% of them have same or better quality than traditional PSTN calls. However, notice that the remaining 15% have fair quality according to the eMOS ranking.

Quite surprisingly, no significant difference can be observed by distinguishing between either source node location, or network access technology. Therefore, in the following, we will try to investigate which network layer index has the largest impact on the eMOS evaluation: specifically, our aim is to isolate and rank the factors that come into play in quantifying the VoIP QoS.

B. Network-centric measurements

As a first cause of eMOS impairment we consider the RTT, although, from the top plot of Fig. 8, we expect its impact on eMOS to be minimal. Indeed, due to the propagation delay properties, we expect the RTT to grow with the geographical distance and to shrink as the access technology bandwidth increase. Fig. 9 reports the RTT CDF faced by different source cities on the top plot and considering different access technology on the bottom plot. As expected, the geographical distance between the source and the gateway has a large

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2007 proceedings.

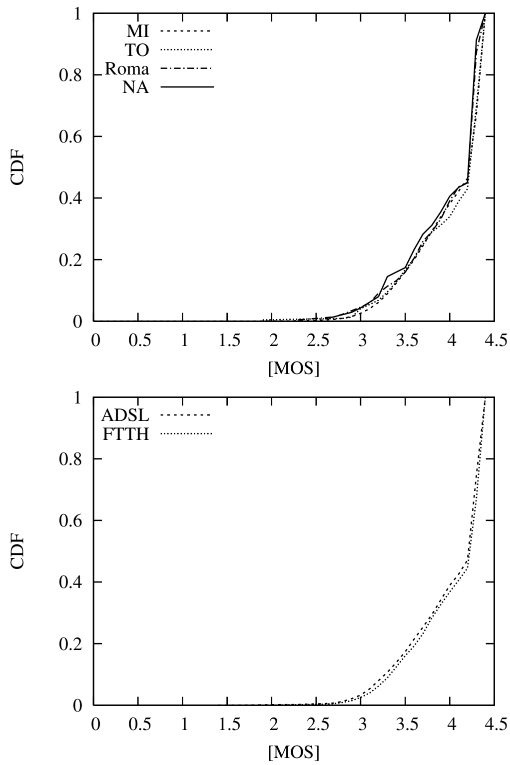

Fig. 8. CDF of eMOS from different cities on the top, and versus network access on the bottom.

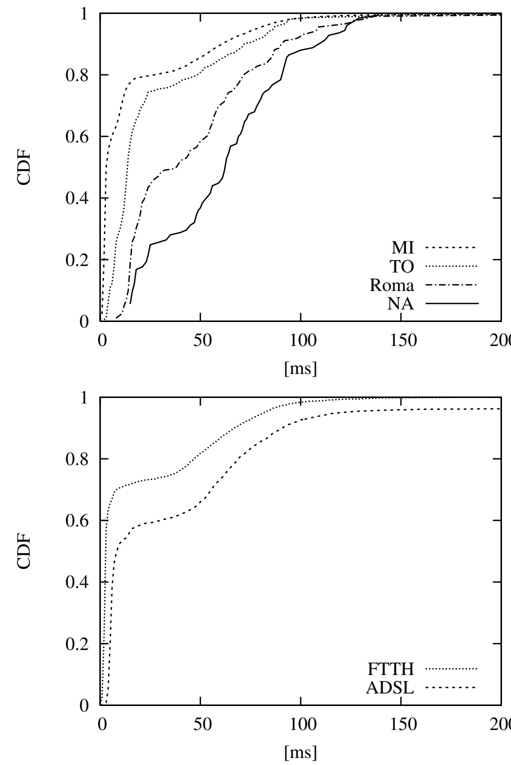

Fig. 9. CDF of RTT from different Cities on the top, and versus network access on the bottom.

impact on the RTT. Indeed, Milan and its hinterland, which correspond to a 20km radius area, show an average RTT of 19.7ms, while calls from Turin, about 140km far from Milan, have an average RTT equal to 27.9ms. Similarly, Rome, 570km far from Milan, presents average RTT of 48.1ms, and finally Naples, 770km far from Milan, exhibits an average RTT equal to 60.4ms.

Considering the different access technologies offered to end-users, it can be observed that Ethernet-based FTTH solution guarantees smaller RTT values (27.3ms on average) compared to xDSL (36.4ms on average). The larger RTT suffered by xDSL users is due to the lower upstream bandwidth, and to possible higher access delay due to ATM framing and bridging to Ethernet adopted in the backbone.

Since all measurements present RTT values smaller than 200ms for more than 97% of calls, we cannot consider the RTT as a major impairment of VoIP call quality.

Jitter may also affect the quality of VoIP calls. However its effect is not directly accounted in the eMOS, since jitter affects the packet loss ratio causing late packets to be dropped by the play-out buffer at the receiver. Therefore, its effect is very difficult to quantify, because it depends on the strategy and settings (e.g., playout buffer length) adopted by the receiver. Since it is not possible to obtain such information from purely passive monitoring, and given that an heterogeneous set of HAG devices are deployed in the FastWeb network, we are forced to not consider jitter as possible cause of packet loss. In order to verify this assumption, Fig. 10 plots the CDF of the measured jitter with the usual convention of reporting curves relative to different cities on top plot, and considering access technologies on bottom plot. Also in this case it is possible to notice that jitter increases with the geographical distance from the gateway probe. Moreover, correlating the jitter with network path length obtained from IP TTL measurements, it is possible to notice that Rome exhibits the largest jitter since nodes are on average 5.5 hops far from Milan, versus 4.75 hops experienced by Naples users, 4.2 and 3.6 hops by Turin and Milan users.

Considering access technology (bottom plot), it is possible to notice that xDSL solution exhibits larger jitter, due to additional encapsulation and DSLAM node presence. Notice that, in all considered cases, jitter suffered by flows is smaller than 15ms, therefore confirming our assumption that jitter has no or little impact on the eMOS evaluation. Indeed, even considering a play-out buffer of a single packet, more than 99% of flows exhibits a jitter smaller than a nominal inter-packet-gap, i.e., 20ms.

Finally, the last parameter that can affect voice quality is the loss probability P_l, whose measurement is reported in Fig. 11 for different cities. Two considerations hold: first, while the average loss probability is equal to 2.8%, the CDF shows that almost 60% of calls did not suffer any packet loss, while the remaining 40% of calls exhibits large dropping probability, up to 20%. Second, almost no difference is observed classifying according to node location. This holds true also comparing FTTH and xDSL access technology. Moreover, looking at

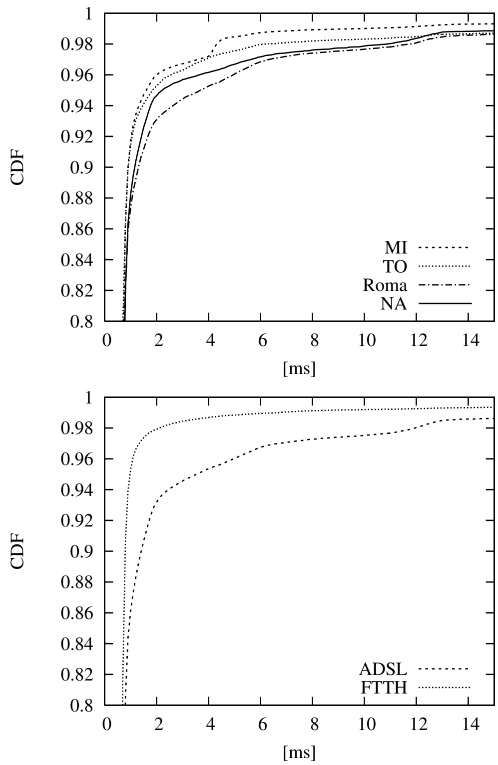

Fig. 10. CDF of jitter from different Cities on the top, and versus network access on the bottom.

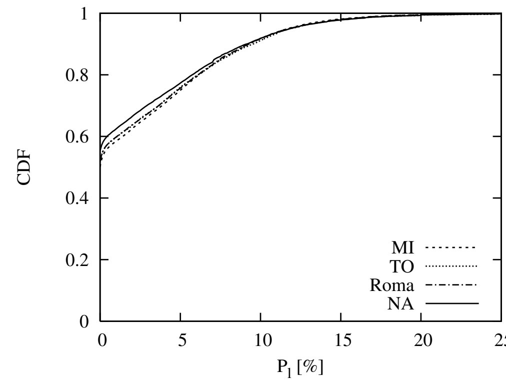

Fig. 11. CDF of packet loss probability from different Cities.

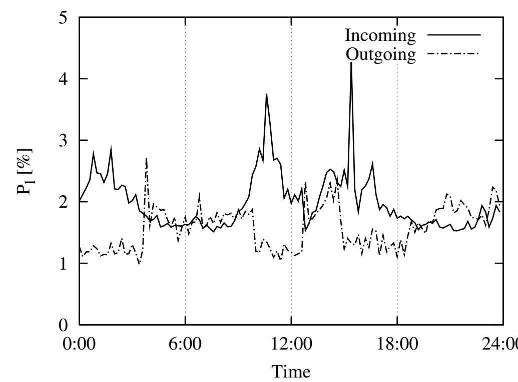

Fig. 12. Average packet loss probability versus time. Milan Gateway Probe.

Fig.12 which reports the average loss probability versus time in a 24h time frame, we note that very little correlation with the actual network load is present (see Fig.5). This is in accordance to the findings in [17], in which authors inspected the loss process observed by TCP flows, and concluded that it is not correlated with the actual network load (provided that no bottleneck is in place at the measuring point).

Given that VoIP and data traffic are carried in the FastWeb network without any QoS differentiation, the loss process suffered by VoIP packets is the same as the one suffered by data packets. Though this does not *compromize* the VoIP QoS on an absolute scale, as its eMOS is always in the expected range, neverthless we are allowed to conclude that the undifferentiated aggregation of voice and data traffic plays an important role in determining the VoIP call quality. Moreover, since we showed that neither RTT nor jitter have a significant impact on the eMOS evaluation in this setup, we conclude that the loss probability is the only network layer impairment affecting the VoIP quality.

V. Conclusions

In this paper we presented an extensive measurement campaign focusing on VoIP traffic characterization. We adopted the eMOS model to compare the quality of VoIP to traditional PSTN phone calls, and investigated the impact of network layer indexes over the eMOS, trying to highlight possible impacts of network topology and adopted access technologies. Measurement results on the FastWeb backbone and PoP infrastructure show that VoIP and network technologies are mature to be deployed by large ISPs, opening the path to the convergence toward a single multi-service network.

Moreover, in the FastWeb solution, measurement results highlighted that only the packet loss probability significantly affected the quality of VoIP perceived by users, while neither delay, nor jitter have a large impact to this end. Neverthless, the presented measurement and eMOS evaluation testify that, for the loss rates we observed, phone-call QoS always falls in the expected range.

Finally, all the algorithms and tools used to obtain the results presented in this paper are made available to the research community via open-source licensing, which we hope will allow other researchers and network operators to contribute to the understanding of multimedia transmission over the Internet.

Acknowledgments

This work was funded by the European Community "EuroNGI" Network of Excellence, and by the Italian Ministry of University, Education, and Research (MIUR) under the PRIN MIMOSA. We would like to thank FastWeb for allowing us to monitor some of their backbone links, and for answering to all our questions.

APPENDIX

To estimate the VoIP call quality, we use the E-model, a computational model standardized by ITU-T through recommendation G.107 [11] that predicts the subjective quality of packetized voice. The outcome is a single rating R (form 0 to 100) which can be further translated into a *Mean Opinion Score* (MOS), in the [0:5] range.

The model is based on the principle that the perceived effect of different impairments is additive, when converted to the appropriate psycho-acoustic scale.

$$R = R_0 - I_s - I_d - I_e + A$$

In this formula the effect of the network is hidden in I_d (delay impairment factor) and in I_e (loss impairment factor). Instead R_0 (effect of noise) and I_s (accounting for loud calls and quantization) are terms intrinsic to the voice signal itself.

In the original scenario, the E-model refers to a different kind of network as the one of a VoIP network. Therefore some adaptations are needed as suggested in [10].

The I_d depends on: i) T_a the absolute one-way delay, ii) T the average one way delay from the receiver side to the first source of echo, and iii) T_r the average round trip delay in the 4-wire loop. However some simplifications can be made. First, all the different measurement points of delay in a VoIP system without switched network networking collapses into a single pair of points such that:

$$T_a = T = T_r/2 = d$$

Second we approximate the one way delay d (not estimable from passive network monitoring) with half the round-trip time:

$$d = RTT/2$$

This value is than increased by the delay due to the Codec used (G.711-A in our case). Due to the lake of information, the dejitter buffer delay is not considered.

I_e is computed from the type of Codec and the packet loss ratio e using a curve fitting method from the data presented in [15].

For all the terms different from the above we took the default values suggested in [11].

Notice that due to complexity of the eMOS evaluation, and the lack of information that is possible to gather doing only passive traffic monitoring, the estimated value of the MOS score is to be considered as an coarse indication of the call quality.

REFERENCES

[1] ITU-T Recommandation H.323, "Packet-based multimedia communications systems", http://www.itu.org, February 1998.

[2] J. Rosenberg and oth., "RFC 3261, SIP: Session Initiation Protocol", http://rfc.net/rfc3261.html, July 2003.

[3] H. Schulzrinne and oth., "RFC 3550, RTP: A Transport Protocol for Real-Time Applications", http://rfc.net/rfc3550.html, June 2002.

[4] R.V. Cox, P. Kroon, "Low bit-rate speech coders for multimedia communication", *IEEE Communications Magazine*, vol. 34, no. 12, pp. 34-41, December 1996.

[5] R. Caceres, P. Danzig, S. Jamin, and D. Mitzel, "Characteristics of Wide-Area TCP/IP Conversations", *ACM SIGCOMM*, Zurich, SW, September 1991.

[6] A.P. Markopoulou, F.A. Tobagi, M.J. Karam, "Assessment of VoIP Quality over Internet Backbones", *IEEE Infocom*, New York, NY, June 2002.

[7] W. Jiang, K. Koguchi, H. Schulzrinne, "QoS Evaluation of VoIP End-points", *ACM SIGCOMM*, Zurich, SW, September 1991.

[8] P. Calyam, M. Sridharan, W. Mandrawa, P. Schopis, "Performance Measurement and Analysis of H.323 Traffic", *Passive and Active Measurement Workshop (PAM)*, April 2004.

[9] A.E. Conway, "A passive method for monitoring voice-over-IP call quality with ITU-T objective speech quality measurement methods", *IEEE ICC*, New York, NY, April 2002.

[10] R.G. Cole, J. Rosenbluth, "Voice over IP performance monitoring", *Journal on Computer Communication Review*, vol. 31, no. 2, pp. 9-24, April 2001.

[11] ITU-T Recommandation G.107, "The E-model, a computational model for use in transmission planning", http://www.itu.org, March 2003.

[12] L. Carvalho, E. Mota, R. Aguiar, A.F. Lima, J.N. de Souza, A. Barreto "An E-Model Implementation for Speech Quality Evaluation in VoIP Systems", *IEEE Infocom*, New York, NY, June 2002.

[13] "FastWeb Company Information", http://company.fastweb.it, 2006.

[14] M. Mellia, D. Rossi, "TCP Statistic and Analysis Tool", http://tstat.tlc.polito.it, 2006.

[15] ITU-T Recommandation G.113, "Transmission impairments due to speech processing", http://www.itu.org, February 2001.

[16] C.Fraleigh, S.Moon, B.Lyles, C.Cotton, M.Khan, D.Moll, R.Rockell, T.Seely, C.Diot, "Packet-level traffic measurements from the sprint IP backbone." *IEEE Network*,V.17, N.6, pp: 6-16, Nov. 2003.

[17] M.Mellia, M.Meo, L.Muscariello, D.Rossi, "Passive Identification and Analysis of TCP Anomalies", *IEEE ICC*, Istanbul, Turkey, June 2006.