Received February 23, 2018, accepted March 27, 2018, date of publication April 2, 2018, date of current version May 2, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2821701

Experimental Analysis of ViLTE Service

CHRISTIAN CALLEGARI[1,2], ROSARIO GIUSEPPE GARROPPO[1], STEFANO GIORDANO[1], CALOGERO CARLO LABROZZO[1], GREGORIO PROCISSI[1], GIOVANNI MINISSALE[3], AND SIMONE TOPAZZI[3]

[1]Department of Information Engineering, University of Pisa, 56126 Pisa, Italy
[2]RaSS National Laboratory, CNIT, 56124 Pisa, Italy
[3]TIM, Turin, 10148 TO, Italy

Corresponding author: Christian Callegari (christian.callegari@cnit.it)

ABSTRACT Long-term evolution (LTE) is a broadband wireless cellular system currently available in about 200 countries. From the mobile network operator point of view, one of the most appealing characteristics of LTE is the possibility of providing the users with mobile broadband services like voice over LTE and video over LTE (ViLTE), with strict quality of service (QoS) guarantees. The success of such services, which usually require a subscription fee, is obviously tightly bound to the users' quality of experience (QoE), that must be significantly better than the quality perceived in case of free services. This paper attempts to evaluate the perceived quality of the ViLTE service. In more detail, we present the results of an experimental test–bed realised at one of the most popular italian MNO, aimed at evaluating the impact of several network layer QoS parameters on the QoE of the ViLTE service.

INDEX TERMS Video over LTE (ViLTE), quality of experience (QoE), LTE, packet loss (PL), packet delay variation (PDV).

I. INTRODUCTION

As of the end of January 2017, Long-Term Evolution (LTE) has been launched in 196 countries, with about 1683 billions of subscriptions worldwide [1]. The high bandwidth offered by LTE permits to offer enhanced services and to improve the quality of experience (QoE) of Over The Top (OTT) services, such as Facebook, YouTube, Video on Demand.

In our previous work [2], we reported an analysis of the applications and services running on top of the LTE system a few months after LTE itself was launched by one of the most popular Italian Mobile Network Operator (MNO). The results evidenced the relevance of the video traffic associated with social networks (such as Facebook), as well as Content Delivery Networks (such as Akamai), and YouTube.

To date, LTE is the most advanced technology deployed to meet the increasing demand for mobile broadband services with Quality of Service (QoS) guarantee. QoS policies based on bearers and QoS Class Identifier (QCI) are managed by MNOs to efficiently deliver acceptable service levels to the end-users [3]. QoS policies in LTE are substantially based on measurable parameters called Key Performance Indicators (KPIs), namely bandwidth, delay, packet delay variation (or jitter), packet loss rate, data rate, and priority. Such QoS native features are crucial for an efficient network management of both data and voice services. In addition, the MNO–managed class of services prove to be more predictable and controllable than the OTT-managed ones [4].

In this field, the LTE system developed its own IP Multimedia Subsystem (IMS) platform to support professional voice and video services with QoS guarantees. Indeed, these services are appealing for their potential to increase the revenue of the MNOs. This fact has stimulated the introduction of Voice over LTE (VoLTE) services, which improve the quality of the transmitted speech through the recommended use of Adaptive Multi-Rate Wideband codec (AMR-WB), also known as HD Voice [5]. At the end of January 2017, 165 MNOs have already invested in VoLTE in 73 different countries, including 102 operators with commercial offers of HD voice services in 54 countries [6]. The evolution of VoLTE, named ViLTE (Video over LTE), also supports a high quality video channel and has been officially launched in less than ten countries, albeit it is currently under test in many others. Nonetheless, the evolution from VoLTE to ViLTE is not straightforward and needs to address a few additional challenges [7].

This paper provides its contribution in this direction by presenting the results of an experimental test-bed realised at one of most popular Italian MNO right before the launch of the ViLTE services. The main objective of this study is to evaluate the impact of several network layer QoS parameters (i.e., packet delay variation – PDV – and packet loss – PL) on the QoE of the ViLTE service. In summary, the main contributions of the paper are:

- realisation of an experimental LTE test-bed at one of most popular Italian MNO;
- evaluation of the QoE of the ViLTE service according to the reference standard [8] under different configuration of the QoS parameters;
- evaluation of a "common" prediction model over the acquired data.

The rest of the paper is organised as follows. Section II gives a brief summary of the related literature and works, while Section III presents the background on the ViLTE service architecture and on the methodology for QoE estimation. Section IV describes the test-bed settings, while Section V presents the analysis of the experimental results and Section VI describes the prediction model and its performance assessment. Finally, Section VII draws the concluding remarks.

II. RELATED WORK

The analysis of video services in cellular systems has received a lot of attention in recent years, albeit most of the research is focused on the evaluation of the QoE of video services offered by OTT. Furthermore, there are also several works that discuss the relationship between QoS and QoE [9], [10]. Focusing on the mobile networks, Vizzarri and Davide [4] study the procedure of QoE metrics estimation from the QoS KPIs measured at network level, focusing on YouTube over the LTE system. In more detail, they apply linear modelling and conclude that the performance mostly depends on the service management approach: when the MNO is able to perform an end-to-end management of the service delivered to the end user, the performance is good, while linear modelling provides unsatisfactory results in case of service managed by the OTT. Amour *et al.* [11] present a test-bed used for collecting subjective QoE scores given by mobile users of YouTube service through a crowd-sourcing approach. Both works differ from ours since they focus on the QoE of video streaming applications offered by OTT, such as YouTube, whereas we address the evaluation of the QoE of an MNO video call service, i.e. ViLTE.

Moreover, most of the previous research refers to MPEG codec, while the ViLTE service specifications [12] define, as mandatory, the H.264 Constrained Baseline Profile (CBP) codec. Regarding H.264 codec, Wang *et al.* [13] present a simulation study carried out with NS-2. The study aims at finding a correlation between QoS and QoE for HD video streaming services over wireless networks. Khan *et al.* [14] investigate a novel content–based, non–intrusive QoE prediction model for low bitrate and resolution (QCIF) H.264 encoded videos. They illustrate the application of the proposed model to video quality adaptation over Universal Mobile Telecommunication Systems (UMTS) networks. The model predicts the mean opinion score (MOS) given the observations of four key parameters: QoE-content type, sender bitrate, block error rate, and mean burst length. The model is described as lightweight and can be implemented for real-time monitoring of the user-perceived quality. QoE–QoS correlation is also addressed in [15], where a survey on machine learning techniques is provided. QoE selected issues in video transmission are also thoroughly addressed in a very recent survey [16] appeared in the first quarter of year 2017. The paper extensively reports on QoE modeling, QoE assessment with subjective tests and objective QoE monitoring, as well as QoE management of video transmission over different types of networks.

In the recent paper Zheng *et al.* [17] report on a database containing subjective assessment scores and the corresponding QoS parameters of 70 video test sequences encoded with H.264, which are corrupted when transmitted over a wireless 3G LTE network simulator. The reference Mean Opinion Score (MOS) values are obtained by using a double stimulus (DS) absolute category rating (ACR) based subjective assessment methodology [8]. Furthermore, they propose a new quality assessment method based on neural networks. The parameters of their method are selected with the aid of simulations, and the QoE prediction accuracy is compared to traditional objective assessment methods. In these works, the authors consider a typical mobile video streaming offered by means of UMTS or LTE networks. Furthermore, the studies have been carried out by means of software simulation of the UMTS/LTE system. On the contrary, our analysis is focused on the Conversational Video Service defined in [12] and the data are obtained by means of experimental tests carried out over an operational network of a popular Italian MNO.

III. THEORETICAL BACKGROUND

A. VILTE SERVICE

The ViLTE services refer to calls with full duplex voice and simplex/full–duplex video media with tight synchronisation between the constituent streams. As opposed to similar OTT services, ViLTE is defined in GSMA specifications by means of a minimum set of mandatory features that wireless devices and the network itself must implement to guarantee an interoperable, high quality IMS-based conversational video service over LTE and/or High-Speed Packet Access (HSPA) radio access. This minimum mandatory set of features is described by the IMS Profile for Conversational Video Service [12].

In addition, a ViLTE session requires the exchange of signalling messages on the IMS platform in order to activate the Dedicated Bearers for the transport of the video and audio RTP traffic. The details of the signalling flows for starting a ViLTE call are described in [5], [12], and [18]. In a nutshell, the User Equipment (UE) connects to the LTE network through an eNodeB and sends a ViLTE request to the IMS platform by using the Session Initiation Protocol (SIP), which is the signalling protocol used by IMS. The IMS network receives the request through the Call Session Control Function (CSCF), and requires the Policy and Charging Rules Function (PCRF) to reserve bearer resources for the involved media. In the Evolved Packet Core (EPC), the CSCF handles the signals from the entire network and retransmits them to the Application Servers (ASs). Hence, the AS executes the service according to the information provided by CSCF. The PCRF is responsible for setting the EPC bearer path, and controls the charging. PCRF informs the Packet Gateway (PGW) and the Serving Gateway (SGW) to allocate the resource and establish the bearers for the ViLTE call. With respect to VoLTE, ViLTE services require the transmission of the UE video capability to the LTE network. It is worth reminding that in a ViLTE call, the PGW and the SGW establish two dedicated bearers for the video call which in most cases are both Guaranteed Bit Rate (GBR). The associated QCI is 1 for the audio traffic and 2 for the video.

A ViLTE session can be established by either adding video to a VoLTE session or by itself. Moreover, in case a GBR bearer used for video fails to establish (or gets lost in the mid-session), the network may still either let the session continue as a voice only call, or terminate the SIP session associated to the GBR bearer.

Hence, to support ViLTE services, the UE must support the ITU-T Recommendation H.264 Constrained Baseline Profile (CBP) Level 1.2, while the support of the H.265 (HEVC) is recommended.

B. QoE ESTIMATION FOR ViLTE SERVICE

The subjective video quality assessment specifications for multimedia applications encompass several experimental methods specifically tailored and validated for different purposes. Such methods include the Absolute Category Rating (ACR), the Absolute Category Rating with Hidden Reference (ACR-HR), Degradation Category Rating (DCR), and the Pair Comparison method (PC) [8]. In all cases, a five–level scale to rate the overall quality (see Table 1) is used. The quality evaluation indicated as MOS is obtained by showing the video "stimulus" to the subjects that rate the magnitude and nature of the sensation caused by the observed video.

TABLE 1. Mean opinion score scale.

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very Annoying

The DCR method differs significantly from the others in that it does not use explicit references. This feature suggests the adoption of DCR when testing the fidelity of transmission with respect to the source signal, as in the evaluation of high quality systems. Instead, the PC method provides high discriminatory power, but the procedure is considerably time expensive.

ACR and ACR-HR are easy and fast to implement and the presentation of the stimuli is similar to that of the common use of the systems. However, ACR-HR has a relevant advantage with respect to ACR in that the perceptual impact of the reference video can be removed from the subjective scores. This, in turn, reduces the impact of scene bias (e.g., viewers liking or disliking a reference video), reference video quality (e.g., small differences in camera quality), and monitor (e.g., professional quality versus consumer grade) upon the final scores.

These features suggest the ACR-HR as the most suitable method for the QoE estimation of video quality in ViLTE services. Indeed, the videoconference clip can be acquired through UEs equipped with different camera and display qualities. The procedure of the ACR-HR method is based on the presentation and rating of the test sequences one at a time. The test procedure also includes a reference version of each test sequence, shown as any other test stimulus, providing a hidden reference condition. In the process of data analysis, a differential quality score (DMOS) is computed between each test sequence and its corresponding (hidden) reference (see [8] for more details on the time pattern for the stimulus presentation).

The DMOS is calculated on a per–subject and per–processed video sequence basis through the following expression:

$$DMOS = MOS - MOS_{ref} + 5 \tag{1}$$

where MOS is the viewer's ACR score of the considered video clip and MOS_{ref} is the mark given to the hidden reference video. A DMOS value of 5 indicates `Excellent` quality while a value of 1 means `Bad` quality. DMOS values greater than 5 (i.e., where the processed sequence is rated of better quality than that of its associated hidden reference sequence) are generally considered as valid.

Finally, it is worth pointing out that the ACR-HR method should only be used with reference videos that have been rated of `Good` or `Excellent` quality by an expert in the field.

IV. EXPERIMENTAL SCENARIO

The experimental setup has been conceived to accomplish the ultimate goal of this study, namely the acquisition of the data about the correlation between QoS and QoE for the ViLTE service. To achieve this target, several aspects must be taken care of in order to avoid measurement polarization. One one hand, all mechanisms and strategies implemented in the ViLTE service must be preserved. Such mechanisms include the ones in charge of managing the control plane functions for sessions setup, the ones devoted to QoS parameters guarantees, and the ones responsible for the provisioning and the configuration of network devices. On the other hand, in order to explore the impact on the QoE of different network conditions, the experimental setup must be able to add and properly tune artificial impairments of QoS parameters.

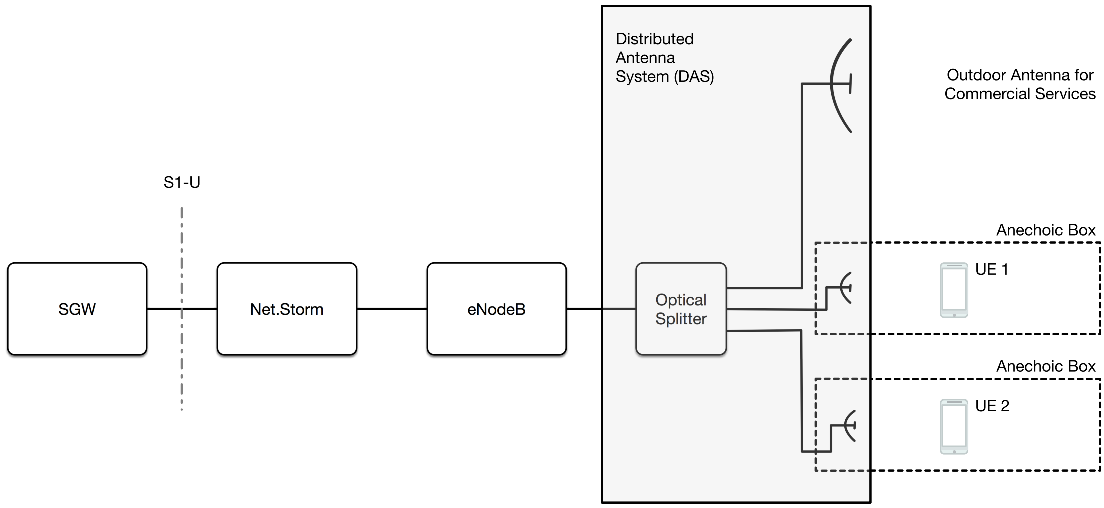

FIGURE 1. Test–bed functional layout.

Given all the above listed constraints, we designed a test–bed that integrates several "ad hoc" network elements necessary for our study into the operational LTE system of a popular Italian MNO.

The integration with the operational LTE system of the MNO allows using the system elements of a ViLTE service provider, i.e. the Core Network, the LTE Radio Access Network, the IP Multimedia Subsystems, and all functions of control plane and management plane used for managing the QoS aspects. Obviously, the results can be influenced by the particular choices of the MNO involved in the study. However, given the relevance of the involved MNO, it can be safely assumed that their specific settings are driven by their long time experience in the different forums (e.g., standardization, network devices manufacturers, service providers, etc.) and in the market. In fact, the different configuration parameters of network and service system should likely be common to most of the MNOs. Hence, the obtained results can be considered quite general and not strictly related to the particular system configuration of the involved MNO.

Beside the operational LTE system, the experimental test–bed additionally includes the two terminals involved in the ViLTE call, a network instrument that adds tunable impairments, and a software tool used to measure QoS performance parameters at the receiver side.

Figure 1 shows the functional layout of the field trial. Once again, we recall that the underlying LTE operational system is provided by the MNO, which allowed us the direct access to some devices of the network for a limited amount of time to carry out our research. In particular, the MNO let us access the S1-U interface to add tunable network impairments on the traffic flow of the observed ViLTE session. More specifically, we used Net.Storm[1] to modify the PL and the PDV in the data plane of the considered IP flows. The measurement approach strictly requires the two UEs involved in the ViLTE session to be connected to the same eNodeB for the whole duration of the experiments, otherwise the Net.Strom would not intercept the examined traffic flow. To solve this issue, we placed each UE into an anechoic box. This way both the UEs are guaranteed yo experience a constant radio quality throughout all experimental sessions, hence preventing from migrating to a different eNodeB. The radio signal is brought into the anechoic boxes by means of an optical splitter. As a result, we implemented a Distributed Antenna System (DAS) that allows to replicate the transmission and reception mechanisms from/to the different antennas and to manage them as a single virtual antenna. In this manner, the signals from the eNodeB to each UEs (and vice-versa) are transmitted through a wired optical link between a dedicated antenna (in the box) and the eNodeB part that normally manages the generation and the elaboration of the LTE radio signals. As a result, the short radio link between the dedicated antenna in the box and the terminal antenna ensures a good radio quality for both UEs. In the outside world, the MNO users are not affected by the experiments as the connection between the eNode and the outdoor antenna of the operational network is maintained. This way, all functions involved in the control of the radio quality and of the reference eNodeB of the two UEs are kept, while maintaining the contention for the radio resources among the ViLTE UEs of our study and the regular MNO customers. As such, this approach also allows to integrate the scheduling and radio resource management strategies used by the MNO for the QoS guarantee of ViLTE services.

In summary, the only approximation that the experimental test-bed introduces concerns the radio quality experienced by both UEs involved in the measurements which is always good since the TX and RX antennas are both placed in the anechoic box. However, this configuration is strictly necessary to maintain full control on the PL and PDV parameters.

[1]Albedo Telecom Net.Storm - Network Impairment Generator

FIGURE 2. **Experimental scenario - terminal side.**

The tests have been carried out by using two last generation smartphones from a popular brand which support the ViLTE services as defined in [12]. As shown in Figure 2, each terminal into its anechoic box is connected by a USB link to a notebook running the TEMS Investigation software.[2] The TEMS allows controlling the terminals, and represents the unique means to access to the terminals since the anechoic boxes cannot be opened during the tests.

In short summary, the TEMS Investigation gives us the following relevant functions:

- remote control of the terminals (i.e. start and tear down of the ViLTE calls between the considered UEs);
- check on the signal quality received by the terminals;
- verification of the correct conclusion of the Attach procedures for the LTE and the IMS system of each terminal, before starting the test. In particular, it allows the identification of the IP address assigned to each terminal, an the information necessary to determine the data flow for adding the network impairments
- data acquisition in terms of PL and PDV at each terminal, during each test.

The considered eNodeB works at 1800 MHz on the operating Band 3, with a channel bandwidth of 1.4 MHz. The eNodeB is physically located close to the test location, as shown by the measured Reference Signal Received Power, which turns out to be −75 dBm. During the tests, the observed Physical Downlink Shared Channel (PDSCH) Signal to Noise Ratio was 40 dB.

A. SELECTING THE VIDEO SEQUENCES

When dealing with Quality of Experience, the selection of the video sequences is a crucial step that must necessarily take into account the feature of the considered service, e.g. video streaming, video on demand, and videoconference. As reported in [8], videos can be classified according to the following two indexes:

- the Spatial Information Index (SI), which represents the spatial information content of the scene. High values of this index indicate a scene with different details to consider;
- the Temporal Information index (TI), which represents the amount of temporal variations in a video sequence. High values indicate scenes with a high variability and sudden changes of environment illumination.

[2] Ascom *TEMS*^{TM} Investigation, version 17.2.3

In our experimental analysis, we have considered two kinds of video: static videos (SV) and dynamic videos (DV). An SV is characterised by a low value of TI and it is associated with a scenario where the environment around the user involved in the ViLTE call is static. This scenario implies that movements in the video are almost exclusively associated with the user's mouth when she talks, and to the slight movements of her head.

Conversely, a DV exhibits a high TI value and it is the typical case of a video call in which the environment around the user changes repeatedly, for example when the user is walking and/or a lot of people and cars move in the background.

Each video sequence lasts 15s and it is characterised by H.264 codec, 640x480 frame size, and frame rate of 15 fps with Constrained Baseline Profile. The values of such parameters come from the default configuration of the smartphones used in the experimental analysis.

B. PROCEDURE AND METHODOLOGY FOR THE EXPERIMENTAL SESSIONS

The experimental sessions have been organised into three distinct phases: the transmission, the estimation of the MoS values, and the definition of a QoE prediction model.

In the first phase, different ViLTE calls involving the two terminals, UE1 and UE2, are generated, each one associated with different settings of the Net.Storm equipment (hence adding different values of PL and PDV to the ViLTE flows). It is clear that the reference sequence used for the estimation of the MOS_{ref} parameter in equation (1) is obtained by setting PL and PDV to zero in Net.Storm. Hence, the different video sequences, as received at destination (i.e., at the UE2 after adding the network impairments) are stored for estimating the associated QoE.

The values of PL and PDV have been selected to cover all of the possible DMOS values, i.e. from 1 to 5. It is worth noticing that such values have been obtained by means of preliminary tests, given that there are no previous studies on the relation between QoE and QoS for ViLTE services.

For each sequence type, namely SV and DV, we carried out two different sets of tests: the first one with 10 different values of PL, and PDV set to 0, and the second one corresponding to 10 different values of PDV, and PL equal to 0.

We did not consider configurations in which both parameters are different than zero. As a first reason, the high number of measurement sessions necessary to consider all meaningful combinations of PL and PDV was not practicably compatible with the time allowed by the MNO for experimentations. In addition, we deem that making a single parameter vary while the other is set to zero allows to clearly evidence possible effects that in the case of combined variation could be easily verified and explained. As an example, in Section V we observed that some PDV values induce a PL in the buffer de–jitter of the devices, which, in turn, worsens the observed QoE.

Furthermore, for each setting pair (PL, PDV), we repeated the test twice. This method should reduce the effects of bias due to short term anomalies in the experimental test-bed.

In each test, the QoS data PL and PDV are measured by the TEMS Investigation. Such values should correspond to the ones set at the Net.Storm, although normally they differ, in practice. Indeed, the TEMS Investigation measures the PL at the application level and includes the effects of the de-jitter buffer, since the measure takes into account the content of RTP packets delivered to the decoder. In particular, the PL is estimated as the percentage of RTP packets sent by UE1 that are not used by UE2 to decode the video. Indeed, packets can be either lost in the network (due to impairments) or at the UE2 (since they are considered useless by the de-jitter buffer). As such, we clearly expect that such an estimation may differ from the value given by Net.Storm.

Instead, the PDV is estimated by TEMS monitoring the RTP timestamp header field and the arrival time of the RTP packet as deduced by the internal clock. In particular, by indicating with S_i the generation time of the packet at the transmitter deduced from the RTP Timestamp header field, and with R_i the arrival time at the receiver and calculated by converting the arrival time obtained from the receiver internal clock in RTP Timestamp unit, the $PDV_{i,j}$ of two successive packets i and j is estimated according to the RFC 3550 as:

$$PDV_{i,j} = (R_j - S_j) - (R_i - S_i) \tag{2}$$

Hence, the PDV values given by TEMS take in account both the PDV introduced by the actual network as well as the one added by Net.Storm which, in turn, only accounts for the PDV added to the ViLTE traffic flows.

Following the procedure of the ACR-HR method sets forth in Section III-B, the second phase addresses the subjective estimation of the DMOS for all of the registered video sequences (i.e., 80 different video sequences, corresponding to a double repetition of 20 different Net.Storm settings for two distinct types of video). At this stage, eighteen (18) individuals were involved in the evaluation.[3] For each sequence, we evaluated the mean value of the DMOS given by all the involved subjects. Overall, at the end of both phases, 80 ViLTE sequences and a set of 80 observations corresponding to the measured triplets $(DMOS, PL, PDV)$ were available for the analysis.

In the third and last phase, the observed triplets were finally used to derive a prediction model.

V. EXPERIMENTAL RESULTS

In this section, we present the experimental results obtained in the different tests and we analyse the impact of the QoS parameters on the estimated QoE.

We start our analysis by measuring the network performance when no impairments are added to the network. To this aim, for each experimental session, we have collected data

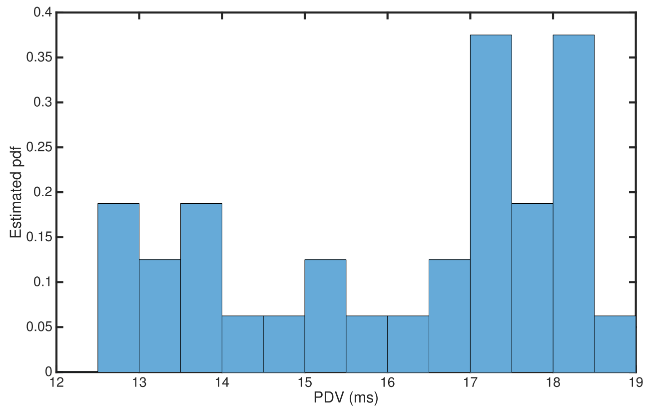

FIGURE 3. Histogram of the average PDV observed before starting each session.

on PL and PDV before when no impairment is set. During this phase we observed no losses, while the estimated pdf of the observed PDV is depicted in Figure 3. The figure shows that the observed average PDV is mainly concentrated in the range [17, 18.5] ms, with a minimum value of 12.58 ms and a maximum of 18.66 ms.

In the next experiments, instead, artificial degradation in terms of PL and PDV is added at the network level. Once again, we recall that the QoS parameters estimated at the application layer by means of the TEMS Investigation normally differ from those set in the Net.Storm, because of the "normal" PDV introduced by the network, which is small (as shown in Figure 3), and of the PL due to the de-jitter buffer of the terminals. Hence, a first set of graphs is devoted to analyse such an aspect.

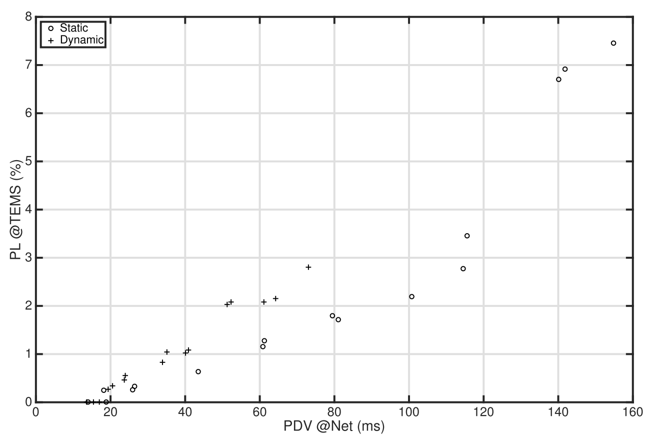

FIGURE 4. PL @ TEMS vs. PDV @ Net.Storm - PL = 0.

Figure 4 shows the PL measured by the TEMS, for different values of PDV in the case of SV and DV. Note that, as already discussed, in this set of tests PL is set to zero at the Net.Storm. Hence, the measured PL is due to the de-jitter buffer only. In case of DV, we have considered PDV values in the range [0, 70] ms, whereas in the static case the upper limit of the PDV is 160 ms. As it will be discussed in the following, this

[3]Note that ITU-T recommends a number of subjects in the range [4, 40] for having statistically significant results

choice is due to the fact that in the dynamic case the DMOS is around 1.8 when PDV is set to 70 ms, which is already a very low value.

Quantitatively, we observe that we do not experience losses when PDV is less than 20 ms, while we observe a packet drop rate lower than 9% for higher PDV values set at the Net.Storm. In particular, in the case of DV, we observe around 3% of losses for the highest PDV (i.e., 70 ms), whereas the PL hits about $7 - 8\%$ for the highest PDV values in the SV case. The analysis of the PDV values measured by the TEMS shows that the difference between the values set at the Net.Storm and the measured ones is in the order of a few milliseconds with a standard deviation of a few units of milliseconds.

A further analysis is devoted to evaluate the difference between the PL set at the Net.Storm and the PL measured by the TEMS, when no PDV is added by Net.Storm. The results of this analysis are summarised in Figure 5, which shows that the measured PL at the TEMS is quite similar to the value set at the Net.Storm. For reading convenience, the bisector in the plot represents the perfect accordance between the values set at Net.Storm and the ones measured by TEMS. It is worth noticing that according to Figure 3 an average PDV in the range [12.58, 18.66]ms is observed in this case, albeit the jitter was set to zero at the Net.Storm.

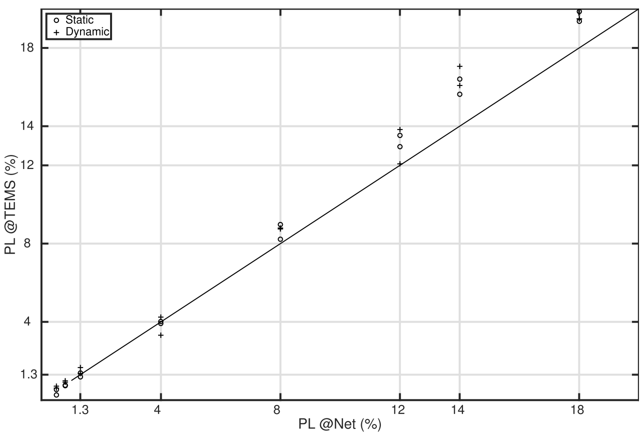

FIGURE 5. PL @ TEMS vs. PL @ Net.Storm - PDV = 0.

A second set of tests is run to analyze the impact of the network impairments on the QoE. To this aim, we first investigate the impact of PL on the DMOS as shown in Figure 6 that depicts the values of the DMOS as a function of PL when PDV is set to 0 at Net.Storm.

From the graph, we can easily deduce that for values of PL lower than 0.5% (extreme left in the graph), the perceived quality is optimal for both SV and DV. On the contrary, when PL increases up to 0.8%, the DMOS is still optimal (about 4) for SV, while the subjects start noticing a slight video deterioration in the DV case (as shown by the `Fair` average DMOS).

By further increasing PL up to 4%, we observe that the DMOS still lays in the `Fair` area for the SV, whereas it drops below 3 for the DV. The `Poor` quality mark (i.e., 2) is then obtained in correspondence of PL set to 8% for DV and to 12% for SV, respectively.

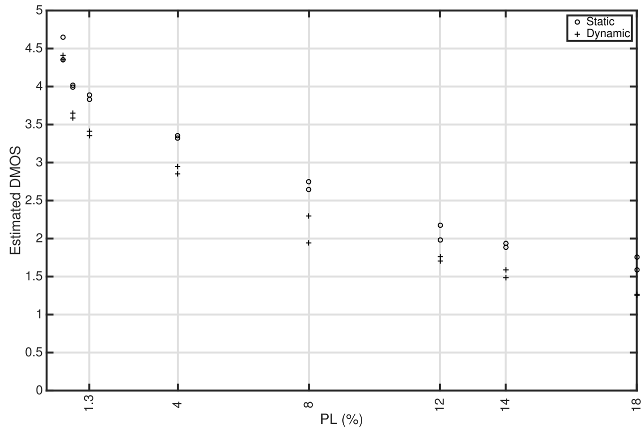

FIGURE 6. PL vs. DMOS - PDV = 0. Values set on the Net.Storm.

In summary, we can conclude that for the same value of PL, the perceived quality is generally lower for DV than for SV. This is due to the fact that the P-Frame generated by the H.264 codec contains more information in the DV case.

Finally, a third analysis has been carried out to evaluate the impact of PDV on DMOS. Figure 7 shows the DMOS values as a function of the PDV. From the graph it clearly emerges that the PDV has generally a stronger impact on the perceived quality of the DV than of the SV. Indeed, the DMOS attains a very low mark (less than 2) for PDV of 70 ms in the DV case, and for PDV of 160 ms in the SV case. Moreover, the perceived quality for the SV is optimal (DMOS greater than 4) with PDV values up to 40 ms and it is still acceptable for PDV values of 60 ms and 80 ms, showing smaller DMOS marks (less than 3) for PDV greater than 100 ms. On the contrary, for the DV case, the subjects start noticing video quality deterioration for a PDV value of 20 ms, while the DMOS goes down to 3 and even less for PDV values equal or greater than 30 ms.

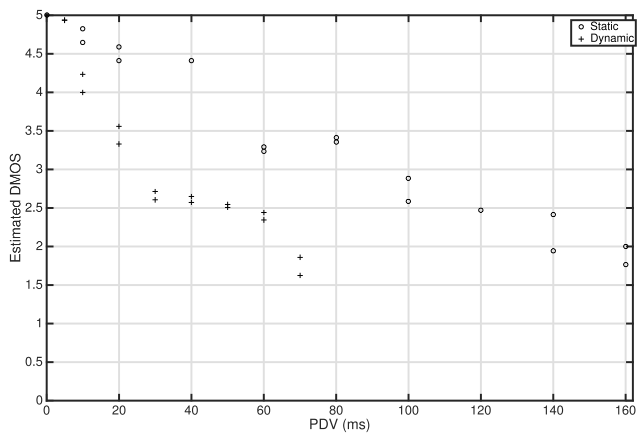

FIGURE 7. PDV vs. MOS - PL = 0. Values set on the Net.Storm.

The above reported impact of the PDV on the perceived quality is likely due to the implementation of the de-jitter buffer in the UEs. Indeed, packets received too late are not delivered to the H.264 decoder. Unfortunately, we could not perform a deeper analysis on this as the implementation of the de-jitter buffer is not detailed in the documentation of the commercial UEs.

Finally, beside the ''agnostic'' analysis of the experimental results, it is really interesting to compare such results to the QCI defined in the standard documents. As already discussed, video transmission in the ViLTE service is associated a QCI of 2, which allows a packet budget delay (PBD) of 150ms and a PL $= 10^{-3}$, see Table 6.1.7 in [3]. In our case, we can easily see that such a percentage of PL (far below the tested values) does not influence the perceived quality. Indeed, DMOS values are optimal for both SV and DV when PL $= 0.5\%$.

Instead, drawing a direct conclusion about delay is not as straightforward. First of all, in our case we should note that the PBD is given by the delay introduced by the network (a few tens of milliseconds in our test-bed) and the delay introduced by Net.Storm. Note that the delay corresponds, on average, to one half of the introduced PDV and, at maximum, to the PDV value. This means that, in the worst case, we can assume that the PDV roughly provides us with (a slightly smaller) PDB value. As a first consideration, we can claim that in the experimental tests the interactivity level of the ViLTE service has always been optimal according to the users (this element is only affected by the packet delay and not by the PDV). Nonetheless, if we consider the impact of PDV on the perceived quality, we can easily observe that the standard QCI values are not sufficient to guarantee an acceptable quality to the users. For example, the DMOS is already smaller than 2 when PDV $= 80$ms for the DV case.

In fact, it is important to point out that such a bad level of perceived quality is not strictly related to delay only, but also to the fact that, as demonstrated in the first set of experiments, high values of PDV induce a not negligible number of losses (mainly due to the de-jitter buffer that discards packets strongly affected by PDV). From an architectural point of view, this is a crucial point, as it also implies that even if not directly specified from the standard documents, MNOs should also consider PDV when delivering ViLTE services. However, under this perspective it is worth mentioning that MNOs do not have control of the full PDV, because it also accounts for the delay introduced by the de-jitter buffer which, in turn, depends on the actual implementation in the users devices.

VI. A PREDICTION MODEL FOR QoE

Based on the experimental investigation reported in the previous sections, the last step of the research was the proposal of a prediction model for the QoE that leverages the knowledge of the set of triplets ($DMOS$, PDV, PL), where PDV and PL refer to the values measured by TEMS. To this aim, we considered a well-known and simple technique. The model is a linear polynomial surface, and the estimated DMOS, denoted as $DMOS_e$, is obtained by a linear combination of terms given by the n-th degree polynomial in the independent variables PDV and PL:

$$DMOS_e = \alpha_0 + \alpha_1 PL + \alpha_2 PDV + \alpha_3 PL^2 + \alpha_4 PDV^2 + \alpha_5 PL * PDV + \dots \quad (3)$$

The model parameters α_i are estimated by means of the linear least squares method, while the degree n has been set by taking into account the best goodness-of-fit measure obtained for different values of n in the range $[1, 9]$. The reference goodness-of-fit measure is the well-known R^2.

At first we attempted to come out with a single model for predicting the correct DMOS value for both the SV and DV video types, but the attained results were not satisfactory. This fact is in accordance with the conclusions drawn in [14], where the content type is pointed out as one of the key parameters influencing the video quality.

Hence, we focused on the definition of two distinct prediction models, one for the DV case and one for the SV case. In both cases, we noticed that a 3-rd degree polynomial provides good prediction accuracy, and that the same model structure applies for both kinds of videos:

$$\begin{aligned} DMOS_e = {} & \alpha_0 + \alpha_1 PL + \alpha_2 PDV \\ & + \alpha_3 PL^2 + \alpha_4 PL * PDV + \alpha_5 PDV^2 \\ & + \alpha_6 PL^3 + \alpha_7 PL^2 * PDV + \alpha_8 PL * PDV^2 \end{aligned} \quad (4)$$

However, the values of α_i in the two cases are quite different, as shown in Table 2.

TABLE 2. Estimated α_i of the two models.

α	SV	DV
α_0	3.207	2.739
α_1	-0.6019	-0.814
α_2	0.0406	-0.3054
α_3	0.4248	-0.5127
α_4	0.7861	0.21
α_5	-0.326	-0.4071
α_6	-0.1514	-0.1757
α_7	-0.3378	-2.356
α_8	-0.1073	-1.755

It is worth noticing that, for the chosen polynomial degree, the model gives an $R^2 = 0.9515$ for the SV case and $R^2 = 0.9787$ for the DV case, hence validating our choice.

A. PERFORMANCE EVALUATION OF THE PREDICTION MODEL

The prediction model was assessed by applying the *leave-one-out* cross-validation approach on the experimentally obtained values reported in section V. The achieved results are almost optimal, with an average mean square error of 0.044 for the DV case, and 0.090 for the SV case.

In more detail, Figure 8 shows the correlation between the Measured DMOS and the predicted DMOS values for the DV data (again, the bisector represents the perfect match) and visually proves the accuracy of the prediction model. In particular, if we adopt the widely used rough classification

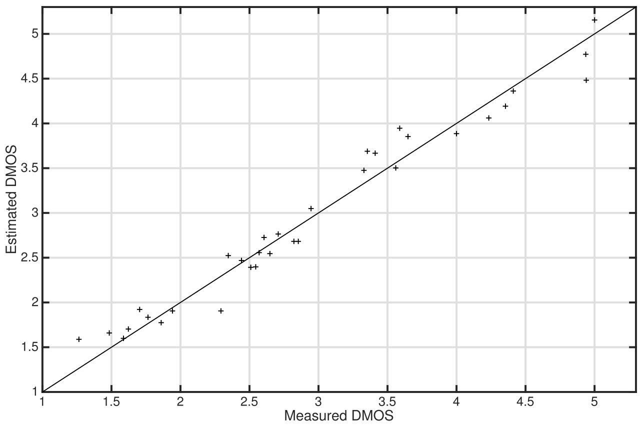

FIGURE 8. **Measured DMOS vs. Estimated DMOS; leave-one-out cross-validation method - Dynamic video.**

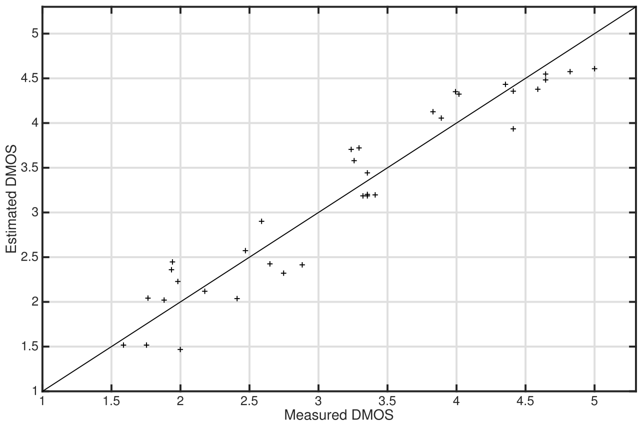

FIGURE 9. **Measured DMOS vs. Estimated DMOS; leave-one-out cross-validation method - Static video.**

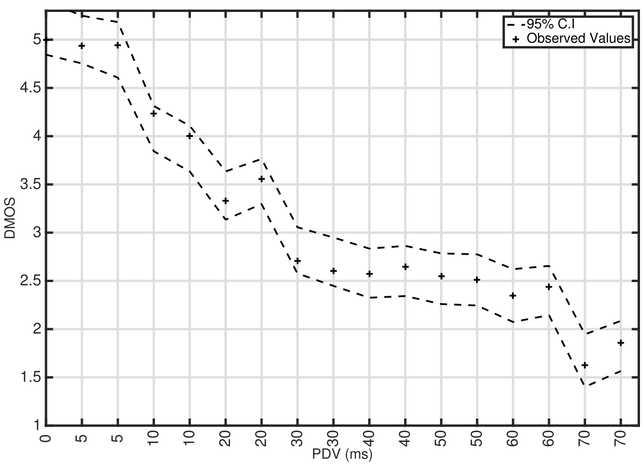

FIGURE 10. **Measured DMOS and 95% C.I. vs. PDV @Net.Storm - TEMS data - PL = 0 @ Net.Storm - Dynamic video.**

of the DMOS values into three clusters, i.e. `Bad` quality for DMOS in the range [0,3), `Fair` in [3,4] and `Good` in (4,5], respectively, we observe that the model always gives a correct results with a single exception. However, notice that even in this specific case, the DMOS value is very close to the border of the `Bad` cluster, i.e. the measured DMOS is equal to 2.9471, whereas the estimated is slightly higher, i.e. 3.0477, but in the `Fair` cluster.

The QoE prediction of static videos, instead, is reported in Figure 9. In this case, we observe a slight worsening of the prediction accuracy, although the "cluster" prediction turns out to be correct in most of the tests, with the exceptions of three `Fair` points that are classified as `Good`, and one `Good` point which is wrongly labeled as `Fair`.

Finally, to further report on the accuracy of the prediction models, Figure 10 shows the position of the measured DMOS values with respect to the 95% C.I. estimated with the prediction model. The results clearly prove that the prediction model is able to provide a reliable estimation of the perceived quality, given that the measured DMOSs always lay within the C.I. bounds.

VII. CONCLUSION

The launch of ViLTE services represents one of the most attracting element of LTE for MNOs. However, for this service to be widely accepted by users, network operators must be confident to be able to provide an optimal level of perceived quality, despite possibly sub-optimal network QoS parameters.

For this reason, in this paper we deployed an experimental LTE test-bed at one of the most popular Italian MNO, with the aim of investigating the impact of PL and PDV on the ViLTE QoE. Interestingly, from the experimental tests carried out by complying the standard procedure [8], it comes out that the limits imposed by the QCIs definition may not be sufficient to guarantee a satisfactory QoE to the users. Indeed, it turns out that the perceived quality is strongly affected by the PDV, which is not considered in the QCI definition. Notice, however, that such a parameter cannot be fully controlled by the MNO, being also dependent on the de-jitter buffer which – in turn – is implemented in the users device by UE manufacturers.

Finally, the paper also analyses the accuracy of a prediction model to estimate the DMOS, given the knowledge of PDV and PL. The experimental results proved the accuracy of the analysed model based on a well-known and simple technique. Despite its simplicity, the model allows to correctly classify the perceived QoE of users in the vast majority of cases.

REFERENCES

[1] G. G. Mobile Suppliers Association. (2017). *LTE-Advanced Carrier Aggregation Deployments: Peak Speeds Report*. [Online]. Available: http://gsacom.com/paper/lte-advanced-carrier-aggregation-deployments-peak-speeds-report/

[2] G. Foddis, R. G. Garroppo, S. Giordano, G. Procissi, S. Roma, and S. Topazzi, "LTE traffic analysis and application behavior characterization," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, 2014, pp. 1–5.

[3] *Policy and Charging Control Architecture; Release 12*, document 3GPP TS 23.203 version 12.6.0, 3GPP, 2014.

[4] A. Vizzarri and F. Davide, "QoE QoS mapping for YouTube services over LTE network," in *Proc. 14th ACM Int. Symp. Mobility Manage. Wireless Access, Co-Located MSWiM (MobiWac)*, 2016, pp. 107–114.

[5] *IMS Profile for Voice and SMS—Version 9.0*, document D. IR.92, GSM Association, 2015.
[6] G. G. Mobile Suppliers Association. (2017). *GSA Status Update: VoLTE, ViLTE, EVS*. [Online]. Available: http://gsacom.com/paper/gsa-status-update-volte-vilte-evs/
[7] Y. C. Sung and Y.-H. Yang, "Challenges from voice-over-LTE to video-over-LTE," in *Proc. 16th Asia-Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Sep. 2014, pp. 1–4.
[8] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document I.-T.P.910, ITU-T Recommendation, 2008.
[9] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1126–1165, 2nd Quart., 2015.
[10] R. Huang, X. Wei, Y. Gao, C. Lv, J. Mao, and Q. Bao, "Data-driven QoE prediction for IPTV service," *Comput. Commun.*, vol. 118, pp. 195–204, Mar. 2017.
[11] L. Amour, S. Sami, S. Hoceini, and A. Mellouk, "Building a large dataset for model-based QoE prediction in the mobile environment," in *Proc. 18th ACM Int. Conf. Modeling, Anal. Simulation Wireless Mobile Syst. (MSWiM)*, 2015, pp. 313–317.
[12] *IR.94 IMS Profile for Conversational Video Service—Version 10.0*, GSM Assoc., London, U.K., 2015.
[13] Z. Wang, L. Li, W. Wang, Z. Wan, Y. Fang, and C. Cai, "A study on QoS/QoE correlation model in wireless-network," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2014, pp. 1–6.
[14] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 431–442, Apr. 2012.
[15] S. Aroussi and A. Mellouk, "Survey on machine learning-based QoE-QoS correlation models," in *Proc. Int. Conf. Comput., Manage. Telecommun. (ComManTel)*, Apr. 2014, pp. 200–204.
[16] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 285–302, 1st Quart., 2017.
[17] K. Zheng, X. Zhang, Q. Zheng, W. Xiang, and L. Hanzo, "Quality-of-experience assessment and its application to video services in LTE networks," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 70–78, Feb. 2015.
[18] *Signalling Flows for the IP Multimedia Call Control Based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3*, document 3GPP TS 24.228 V5.15.0, 3GPP, 2006.

CHRISTIAN CALLEGARI received the Laurea degree *(cum laude)* in telecommunication engineering from the University of Pisa in 2004, discussing a thesis on simulative analysis of RSVP-TE, and evaluation of end-to-end rerouting techniques in MPLS networks, and the Ph.D. degree in information engineering from the Department of Information Engineering, University of Pisa, in 2008. In 2005, he obtained the qualification to practice the profession of an Engineer. He is currently a Post-Doctoral Fellow with the Telecommunication Network Research Group, Department of Information Engineering, University of Pisa. His research and professional areas of interest are network security, traffic classification, traffic engineering, MPLS architecture, and network simulation. He became a member of the IEEE Communication Society in 2005.

ROSARIO GIUSEPPE GARROPPO received the M.S. (Laurea) degree *(cum laude)* in telecommunications engineering and the Ph.D. (Dottorato di Ricerca) degree in information engineering (ingegneria dell'informazione) from the University of Pisa in 1995 and 1999, respectively. He is currently an Assistant Professor with the Dipartimento di Ingegneria dell'Informazione, University of Pisa. He has published over 100 peer-reviewed papers in international journals and conference proceedings. His expertise is on networking, and his main research activities are focused on experimental measurements and traffic modeling in broadband and wireless networks, MoIP systems, traffic control techniques for multimedia services in wireless networks, network optimization, and green networking. He served as a technical program committee member for several international conferences on wireless networks, and a referee for several international journals and conferences. He received the Best Paper Award from the 4th International Workshop on Green Communications in 2011.

STEFANO GIORDANO received the Laurea degree *(cum laude)* in electronics engineering and the Ph.D. degree in information engineering from the University of Pisa, in 1990 and 1994, respectively.

He has been involved in telecommunication networks with Consorzio Pisa Ricerche since 1990, where he has been participating and coordinating several research activities. Since 2001, he has been an Associate Professor with the Telecommunication Networks Group, Department of Information Engineering, University of Pisa, where he has also been a Lecturer of telecommunication networks and the design and simulation of telecommunication networks. His research and professional areas of interest are broadband communications, telecommunication networks analysis and design, simulation of communication networks and systems, and multimedia communications.

CALOGERO CARLO LABROZZO received the master's degree in telecommunication engineering from the University of Pisa in 2016. In 2016, he was an Intern at TIM, during which he was involved in end-to-end testing and performance evaluation on LTE/IMS mobile networks and mobile handsets, impairment injection, protocol analysis, and troubleshooting. Since 2017, he has been an NOC Engineer with Alten Italia.

GREGORIO PROCISSI was with the Basic Research Institute in the Mathematical Sciences and also with Hewlett-Packard Labs, Bristol, U.K., in 1998. From 2000 to 2001, he was a Visiting Scholar with the Computer Science Department, University of California at Los Angeles, Los Angeles. He has been an Assistant Professor with the Department of Information Engineering, University of Pisa, since 2005. His research activities, published over 70 papers, including design and performance evaluation of high-speed networks, traffic modeling and control in IP networks, passive and active traffic measurements, deterministic, and stochastic techniques for deep packet inspection. He has been involved in several national and European projects on network monitoring, including PRISM and DEMONS.

GIOVANNI MINISSALE has been an Electronics Engineer with TIM since 2002. He was involved in testing HW/SW tests for GSM/UMTS radio access systems and testing new technologies from HSPA+, femto, and LTE. He has participated in various international networking and network optimization campaigns in Greece from 2002 to 2003, Chile in 2008, Cuba in 2010, and strategic consultancy, Jakarta, in 2010. Within the Radio Lab area, he has been responsible for the Innovative Testing and Network Diagnostics Support Program with the main purpose of managing the testing of innovative radio access solutions for future generation networks, trials, and projects special (e.g., TAV) and support territorial areas.

He is currently a Project Manager for mobile device certification testing and developing, with the main purpose of validating TIM mobile devices to be inserted into TIM portfolio, and support developing of LTE-A, VoLTE, and Internet of Things modules toward 5G.

SIMONE TOPAZZI joined the Company (former CSELT) to provide consulting support to the Group's Foreign Affiliates, in view of the acquisition of new radio licenses, network start-up support, and development of network evolution/ innovation technology plans in 1998. He has been with the laboratory and testing topics since 2008. He is currently the Head of the TIM Mobile Solution Verification Department, where he was responsible for the testing, validation, and acceptance of mobile terminals, radio access and core technologies, and management systems for the mobile network, and also a Telecommunications Engineer.