Assessing QoE-driven management policies for VoIP and Video Streaming service provisioning

D. Ciambrone*, S. Tennina*, M. Boschi*, D. Tsolkas[†] and L. Pomante[‡]

*WEST Aquila S.r.l., L'Aquila, Italy

{daniele.ciambrone, tennina} at westaquila.com

[†]Department of Informatics and Telecommunications, University of Athens, Greece,

dtsolkas at di.uoa.gr

[‡]Center of Excellence DEWS, University of L'Aquila, Italy

luigi.pomante at univaq.it

Abstract—In the frame of the CASPER project, we envision the performance evaluation of the new user-centric network and service management for future mobile networks. This paper extends our previous contribution on this subject by detailing a novel system level simulator built upon SimuLTE/OMNeT++ to assess quality of experience (QoE)-aware management policies in real LTE networking conditions with Voice over IP and video streaming traffic patterns. The results demonstrate that an accurate choice of the design parameters both at the application layer and at the network layer is paramount to achieve acceptable QoE levels over the time.

Index Terms—QoE, VoIP, Video Streaming, LTE, System Level Simulation, Packet Loss Rate, End-to-End Delay.

I. INTRODUCTION

Quality of Experience (QoE) is a recently proposed acronym to evolve from the well-known Quality of Service (QoS) to express the perception of the quality of both network and service from the end-user perspective [1]. QoE quantifies end-users (e.g., video consumer's) satisfaction level. It is indirectly correlated with user's engagement with the service and is a crucial factor for enabling advanced customer experience management (CEM). The primary objective of CASPER project [2] is to combine academic and industrial forces towards leveraging the expected benefits of QoE exploitation in 5G networks.

CASPER aims at the definition of an integrated middleware architecture, following industrial requirements and recent standardisation activity. The user-centric logic of the CASPER project is based on an in-depth QoE analysis and adopts mainly results from the available literature (e.g., [3]), combined with the flexibility and reconfigurability provided by the software-defined networking (SDN) [4] and network function virtualisation (NFV) [5] technologies.

For a beneficial intervention of QoE in network and service management in the 5G era three interlinked modules are proposed [6]: (i) Reliable and passive QoE monitoring, including adaptive and objective QoE estimation mechanisms; (ii) Robust and real-time QoE-driven service management and CEM policies; (iii) Efficient QoE controller to realise the interface between the middleware and the network functions.

In this paper the objective is to present our system level simulator built upon SimuLTE/OMNeT++ [7], [8] to reproduce the CASPER reference scenarios and evaluate the effects

Fig. 1. Video streaming/Online video consumption.

of monitoring and control (i.e., management) policies over the final QoE perceived by the end users. In particular, this paper extends a previous work we presented in [9] with new details of the simulation environment, the use case scenarios, and the QoE management policy.

Consequently, the remainder of the paper is organised as follows. Section II briefly recalls the scenarios addressed in the CASPER project and meaningful for the work of this paper, while Section III details the implementation we did of those scenarios into the simulation framework. After describing the simulation setup, Section IV discusses the impact of the management policies on the QoE for both VoIP and Video Streaming services. Finally, Section V concludes the paper.

II. SCENARIO DEFINITIONS

For this paper, the work has been focused on variants of the second and third scenario of CASPER project, because more thorough and comprehensive as compared to the first one. Accordingly, just these are recalled here, while the full description is available in [9].

A. Multimedia Content Delivery from a Service Provider

This scenario mainly concerns cases of video content delivery, namely video streaming. The prevailing paradigm nowadays is that the Over The Top (OTT) parties provide access to video streaming services, that the users can access

Fig. 2. Skype call or web conference.

only through their MNO or ISP connection¹. This scenario of CASPER is presented in Fig. 1.

The main applications that are going to be studied here are: (i) Non-adaptive video streaming (non-real-time service). (ii) HTTP Adaptive Streaming (non-real-time service). (iii) Live TV / IPTV (real-time service).

The major challenges in this category of scenarios stem from the fact that the video content is usually made available by an OTT provider, whereas it is delivered by a traditional MNO/ISP. Therefore, QoS/QoE guarantees from the OTT side are mostly absent and are mainly oriented to providing high-resolution video content to the viewers. However, the delivery procedure of the content itself is at the discretion of the MNO/ISP. Consequently, if the quality of the offered video is terrible despite the OTT's proactive measures, a closer interaction between MNO/ISP and OTT is needed to improve the system's efficiency from the end user perspective.

B. Real-time Communication among Multiple Parties via an Intermediate Server

This scenario is the most challenging one since it includes an OTT provider, one or more MNO/ISP providers and multiple users. Applications that fall into this category are Skype calls (voice or video), Viber, gaming, and teleconference, among others. The traffic, in this case, is accumulated at a server in possession of the OTT party, passing however through the network infrastructure of one or more MNO/ISPs (intra-operator and inter-operator case, respectively). This scenario is presented in Fig. 2.

The challenges, in this case, include the ones discussed in the previous scenario, as well as the fact that this one involves multimedia content delivery. A major challenging issue from a QoE perspective is also that proper and reliable QoE estimation models for these types of applications are not available yet, and finding appropriate KPIs is not a trivial issue. The CASPER architecture will tackle these challenges by providing interfaces to all involved parties for QoE monitoring as well as control. It will, therefore, serve as a platform that enables interactions among the different stakeholders, with the target of providing user-personalised, application-aware and context-aware QoE improvements.

¹MNOs or ISPs may also provide video streaming services by themselves to their subscribers.

Fig. 3. Implementation of the scenarios of Fig. 1 and Fig. 2 in the chosen simulation environment.

III. IMPLEMENTATION AND SIMULATION

The simulation tool chosen to reproduce the CASPER scenarios, as recalled in the previous section, is SimuLTE [7], which allows reproducing LTE-based wireless and wired network architectures, building on the INET Framework and the OMNeT++ discrete event simulator [8].

To implement the main scenarios in the simulation environment, we have defined in [9] a basic network topology, to simulate them. So, this paper is an evolution of that previous one, since we introduce several updates in the network topology, traffic mechanisms and QoE measurements and management policy, to describe and simulate more complex applications of VoIP and Video Streaming. The reference network topology is shown in Fig. 3.

Without loss of generality, let's start describing the characteristics of the scenario of Fig. 2. It deals with VoIP-based traffic patterns between user equipment (UEs), over a single or multiple MNO/ISP providers and passing through an OTT-server. The implementation of this scenario is shown in Fig. 4.

As already claimed in [9], in the simple approach of a static network topology sketched simulations execute in ideal conditions, where packet loss rates and delays are always low and mainly depends on the propagation conditions in the wireless access portion of the LTE architecture. However, in the real world, network congestions and other impairments happen, which affects the packet loss rate and the end-to-end delay. The objective is to control the simulation parameters so that the evidence about the tradeoff of packet loss rate and end-to-end delay for the perceived QoE at the end-users is given. To do this, we imagine that at a given time $t = T_0 > 0$ an impairment arises in the network, e.g., a network congestion leading to a sudden rise in the packet loss rate: our QoE monitoring entity recognizes it at time $t = T_0 + R$, where $R > 0$ is called *Reaction Time*, and starts taking the appropriate control actions.

Accordingly, in this new implementation, we focused on having multiple parallel flows, i.e., VoIP calls, between pairs of UEs and the new topology is characterised by two further uplink and downlink paths, used as backup paths. We implemented a real-time QoE-driven management policy to trigger a change of path in the network when the QoE level falls below a chosen threshold and also a mechanism to implement

Fig. 4. Implementation of the scenario in Fig. 2.

different priority levels to handle the traffic of the VoIP calls in the OTT server.

More in detail, in this VoIP application, there are five parallel communications between five pairs of *UEs*, which generate and inject packets into the network through the main path, according to the VoIP-based traffic pattern, which includes talk spurts and silences based on a Weibull random distribution. The QoE is continuously monitored by computing the level of the Mean Opinion Score (MOS) as by the E-model [10] and the ITU P.563 [11] for VoIP service and the ITU-T G.1070 [12] for Video Streaming service. A path's change from the main path to the backup path is triggered if the value of MOS persistently falls below a given threshold, i.e., if it holds that $MOS(t) < MOS_{TH}$, for $t \in [T_k, T_{k+countMOS}]$. Basically, we consider a window of the most recent MOS values, and we check if the condition is true for the last *countMOS* values.

For our simulations, the backup path is characterised by a lower packet loss rate, to guarantee a recovery regarding network's efficiency. However, at the same time, it leads to a slightly higher end to end delay, taking into account that the backup path is longer than the main one and is a sub-optimal choice. The combination of these two factors should guarantee an overall improvement on the QoE perceived at the users' side, so that the network impairment is considered resolved at time $t = T_0 + D$, where $D > R$.

Furthermore, in the case of VoIP calls, it has also been implemented a different priority level of the call, which impacts on the queueing mechanism on the OTT server, where the packet of low priority calls are buffered before being forwarded to the destinations, while those belonging to high priority calls are directly forwarded.

Similar considerations can be done for the Video Streaming scenario. As explained in [9], a video streaming service runs at the application layers of a variable number of *UEs* and the OTT server. Each *UE* sends first a request to the server, which then starts transmitting the chunks of the video. The download can be executed simultaneously by one or more users or in different periods. In our case, we have 9 *UEs* involved in a simultaneous download of a video. The implementation of this scenario is shown in Fig. 5.

In terms of implementation details, we kept the same structure previously adopted in [9], but the new changes apply to the modules *VoIPReceiver*, *VoIPSender*, *VoIPServer*

Fig. 5. Implementation of the scenario in Fig. 1.

and their *middleware* modules, for the scenario of Fig. 2, and *VideoStreamCli*, *VideoStreamSvr* and their *middleware* modules, for the scenario of Fig. 2, respectively.

The setup of these reference applications to execute the simulations and obtain results is detailed in next section.

IV. SIMULATION SETUP AND RESULTS

A. Simulation Setup

About the scenario of Fig. 2, the *MOS* expression depends on the average packet loss rate (*PLR*) and end-to-end delay (*E2E*). In our simulated world, these statistics are computed at run-time by the receiver in the C++ code of the *Middleware* module, transparently to the *VoIPReceiver* application module. In more detail, the VoIP traffic has a pattern of random duration of talks and silences periods and these random values are distributed according to a Weibull function. Thus, for the computation of the average *PLR* and *E2E* to be meaningful, we implemented a mechanism that updates counters of packets received and timestamp values into a buffer at the receiver side. Then, every $T_{refresh}$ seconds the statistics are computed, and the buffer is reset. Furthermore, we opted to allow the computation only if a given number of packets is transmitted, to avoid having to inconsistency in the data.

In our simulation framework, we control the statistics of the packet loss rate and end-to-end delay according to given random distributions, i.e., each received packet at the *Middleware* layer is discarded (i.e., lost) if the result of a call to *intuniform(0,100)* is less than the pre-defined $\hat{PLR}$ threshold. Then, if the packet is not discarded, it will be forwarded to the application layer after a random delay, representing the end-to-end delay, resulting from $E2E \sim \mathcal{N}(\mu, \sigma)$, i.e., following a Gaussian Distribution whose average and standard deviation parameters are given. Accordingly, to set an appropriate value for these different distribution, we started by looking at the results achieved in [9]. In particular, Fig. 6 show the distribution of the *MOS* field for the VoIP traffic in the scenario of Fig. 2, against the combinations of $\hat{PLR}$ and μ , for $\sigma = 0.01$ s.

As shown, once a value of MOS_{TH} is decided as a threshold to consider the quality of experience as acceptable, a range of *PLR* and *E2E* is given. For instance, if $MOS_{TH} = 3.5$ and the average *E2E* is around 100 ms, then the *PLR* should remain below 25%.

Consequently, to assess the effects of a real-time policy which will lead to a path's change, we have simulated the trigger of networks' impairments in consecutive events, such

Fig. 6. Average QoE/MOS of VoIP of the scenario of Fig. 2 against the combinations of PLR and μ for $\sigma = 0.01$ s [9].

as introducing an extra end-to-end delay and then a packet loss rate, i.e., $\mu(t) = \mu_1$ for $t \geq T_{e2e}$ and $PLR = P_1$ for $t \geq T_{plr}$, respectively, with $T_{e2e} \leq T_{plr}$. The introduction of a higher PLR determines a decrease of the MOS level which might falls down the reference threshold (MOS_{TH}) and deals with a change of the path. In the new path, as anticipated early in Section III, the packet loss probability reduces to $PLR = P_2$, with $P_2 < P_1$, since the new path is assumed of a higher reliability than the previous one. However, a lower value of PLR corresponds a higher value of $E2E$ since the new path is supposed to be longer (suboptimal) than the previous one: so, $\mu(t) = \mu_2$ for $t \geq T_{pathChange}$, with $\mu_2 > \mu_1$ and $T_{pathChange}$ is the instant when the path changes, which depends on the *Reaction Time* (R). Finally, on the new path, the combination of decreased PLR and increased $E2E$ should determine an increase in the level of MOS again, near or above the threshold (MOS_{TH}).

Furthermore, we assume that at time $t \geq T_{prio}$ one *UE* starts marking its VoIP calls as low priority. This will instruct the OTT server to buffer the packets of this flow. In detail, this means that when a low priority packet is received at the OTT server, it is enqueued and, if not already scheduled, a timer is set equal with the interval set to T_{queue} : at the timeout, all the packets present in the queue are then transmitted. Consequently, a differentiation of flows is shown, whose effect is that the values of MOS for the delayed/buffered traffic are lower than those with high priority. Without loss of generality, we assume $T_{prio} \gg T_{pathChange}$.

Similar considerations hold for the scenario of Fig. 1. The QoE policy for Video Streaming will provide reaction to a change in the MOS value due to an increase in the end-to-end delay and packet loss rate. When the MOS falls persistently below the MOS_{TH} threshold, a path's change is triggered to the backup path where the PLR is decreased, and the $E2E$ is slightly increased. The main difference with the VoIP application lies in the mechanism used to compute the real-time value of MOS during the simulation. Contrarily to the VoIP traffic pattern, the streaming application produces a continuous generation of packets without silence periods.

Fig. 7. Best achieved QoE/MOS for Video Streaming of the scenario of Fig. 1 against the PLR , for different video coding options.

Table I
SIMULATION SETUP PARAMETERS.

Parameter	VoIP	Video Streaming
Simulation Time	100 s	100 s
MOS_{TH}	3.5	3.5
$countMOS$	[10, 50]	[50, 100]
P_1	30%	30%
P_2	5%	[5, 9, 15]%
μ_1	100 ms	100 ms
μ_2	105 ms	105 ms
σ	10 ms	10 ms
T_{e2e}	10 s	10 s
T_{plr}	30 s	20 s
T_{prio}	50 s	N.A.
T_{queue}	[50, 100] ms	N.A.
$T_{refresh}$	1 s	N.A.
Video Size	N.A.	500 kB
Packet Length	N.A.	200 B
Video Encoder	N.A.	H.264
Frame Rate	N.A.	25 fps
Bit Rate	N.A.	5000 bps
Send Interval	N.A.	40 ms

Thus, we opted to avoid a refreshing of the local buffer on the receive side, where the average packet loss rate is computed to estimate the MOS. Instead, we use a circular buffer to compute the MOS value every received packet, after a transient, to accumulate some packets received for data consistency reasons. As for the VoIP case, our study starts from assuming the results of Fig. 7, already presented in [9].

Table I shows the simulation setup: it recaps the main parameters and their values.

B. QoE Results

Assuming to accept a threshold of $MOS_{TH} = 3.5$ for the QoE that is considered as acceptable by end-users for the given application service (i.e., in this case of both VoIP and Video Streaming), and applying the real-time QoE-management policy as previously described, we have obtained the results shown in the Fig. 8 and in Fig. 9, for VoIP and Video Streaming services, respectively.

Fig. 8. Example of a VoIP traffic result for the scenario of Fig. 2

With reference to Fig. 8, it is evident that, after the initial flat trend, where PLR and $E2E$ are set equal to zero, with the introduction of the end-to-end delay at time $T_{e2e} = 10$ s, the MOS slightly degrades, with a semi-linear trend. This is the effect of the increased $E2E$ delay in the MOS formulation, but also the fact that every time we compute the MOS, i.e., every $T_{refresh} = 1$ s, there is also a number of packets that has not been received yet, which contributes to the computation of the PLR . In any case, this double effect produces a slight degradation in the MOS, which settles down just below the value of 4.

Then, at time $T_{plr} = 30$ s, the packet loss probability ($P_1 = 30\%$) is introduced: this implies a vertical fall of the MOS, as a direct consequence of such a high value of PLR , which leads the MOS to fall down the adopted MOS_{TH} threshold. Once the computed MOS keeps below the threshold for $countMOS$ consecutive times, the QoE management policy starts to react and leads to deviate the traffic flow to the more reliable but slightly slower backup path at time $T_{pathChange} = 31.984$ s, i.e., leading to a *Reaction Time* $R = 1.984$ s. This quickly raises the value of MOS, eventually again above the MOS_{TH} threshold. It is worth stressing that the reaction time after $t = T_{plr}$ is correlated with the value of the $countMos$ parameter, which determines the width of the basin where the MOS lies below the threshold. Moreover, in this scenario, at time $T_{prio} = 50$ s, the last flow (Flow number 4 in Fig. 8) starts using low priority traffic, which leads to start of enqueueing packets on the OTT server, and determines a further delay in the end-to-end chain for that flow. This leads to a net decrease of its MOS.

Similar considerations hold for the Video Streaming service of the scenario of Fig. 1. About Fig. 9, the trend for the different UEs is very similar to each other, and with the previous VoIP application case. The simulation setup is as in Table I and linked to the combinations previously described in [9], as reported in Fig. 7. As for the previous case, after introducing the effects of PLR at time $T_{plr} = 20$ s, the MOS value quickly degrades and falls down below the given MOS_{TH} threshold. Consequently, a change of path is triggered at time $T_{pathChange} = 23.4981$ s, i.e., leading to a

Fig. 9. Example of a Video Streaming traffic result of the scenario of Fig. 1

Fig. 10. VoIP traffic result for the scenario of Fig. 2 and combination of parameters in Table I

Reaction Time $R = 3.4981$ s, which is greater than the case of VoIP, mainly because of the higher $countMOS$ value.

Widening the perspective on the overall setup space identified by the combinations of the parameters in Table I, in Fig. 10 and Fig. 11, the results of several simulations are compared to show the effects of the adopted QoE policy for VoIP and Video Streaming scenarios, respectively.

About the VoIP application, Fig. 10 shows the trend of MOS values as a function of time, related to different sets of simulations. More in detail, we have represented the average trend for four flows with high priority for different values of T_{queue} on the server and packet loss probability. Meanwhile dotted curves represent the VoIP flow at low priority, whose packets are buffered on the server and send after a delay. In particular, we can observe how the network reacts to the introduction of the increased packet loss probability by changing the path. The value of $countMos$ determines a greater width of the basin in the graph and the values of the reaction time, as reported in Table II.

Fig. 11. Video Streaming traffic result of the scenario of Fig. 1 and combination of parameters in Table I

Table II
REACTION TIME FOR VOIP APPLICATION

countMOS	Reaction Time
10	1.984 s
50	8.0083 s

As far as the Video Streaming application is concerned, Fig. 11 demonstrated that we had obtained similar results as VoIP. When varying the packet loss probability on the backup path P_2 from 5% to 15%, the resulting MOS values are above, on edge and below the threshold, respectively, and this is in good agreement with Fig. 7. Once again, the value of *countMos* determines a greater width of the basin in the graph and the values of the reaction time, as reported in Table III.

In the big picture, we can observe that the real-time QoE policy has allowed reacting against the network impairments, guaranteeing an acceptable value of MOS near the defined threshold. This is the proof that it is possible to act on further metrics to get even better results and there is space to optimise the parameters of the QoE management policy to boost it up and make it more efficient.

V. CONCLUSION

In this paper, we have given an overview of the system level simulator proposed in the frame of CASPER project. In more detail, we recalled the reference scenarios addressed in the project, and we sketched how they are reproduced within the simulator. Then a detailed performance analysis has been given for the VoIP and the Video Streaming application

Table III
REACTION TIME FOR VIDEO STREAMING APPLICATION

countMOS	Reaction Time
50	3.4981 s
100	6.3238 s

services in the presence of multiple parallel flows, real-time QoE continuous monitoring and traffic differentiation at the OTT server. Our analysis demonstrated that the adoption of QoE-driven management policies which recognise network impairments and change the path where the traffic flows is paramount to maintain the quality of the services above an acceptable threshold.

Our on-going work is revolving around the implementation of different applications than VoIP and video streaming, e.g., the HTTP Adaptive Streaming. In this application, another parameter enters the game: the packet size that can be adjusted to mitigate the effects of variations in the network throughput and jitter. However, this assumes that a different metric for the QoE than the formulations proposed here is adopted, where throughput and jitter appear in the expressions, as, e.g., the metrics proposed in [13].

ACKNOWLEDGEMENT

The authors would like to acknowledge the European Commission project CASPER (Grant number 645393), an H2020-MSCA-RISE action. They also acknowledge Greek Financial Support from the State Scholarships Foundation (IKY) in the context of the action *Excellence for Postdoctoral Researchers NSRF 1st Cycle for 2014-2020* (programme: Human Resources Development, Education and Lifelong learning).

REFERENCES

- [1] R. Stankiewicz, P. Cholda, and A. Jajszczyk, "QoX: What is it really?" *IEEE Communications Magazine*, vol. 49, no. 4, pp. 148–158, April 2011.
- [2] A. Colarieti, A. Marotta, M. Mpervarakis, L. Pomante, and V. Tsolkas, "QoE provisioning over mobile networks: The CASPER perspective," in *2017 IEEE 22nd CAMAD*, June 2017, pp. 1–5.
- [3] D. Tsolkas, E. Liotou, N. Passas, and L. Merakos, "A survey on parametric qoe estimation for popular services," *J. Netw. Comput. Appl.*, vol. 77, no. C, pp. 1–17, Jan. 2017. [Online]. Available: <https://doi.org/10.1016/j.jnca.2016.10.016>
- [4] O. Networking Foundation, "Software-defined networking: The new norm for networks," 04 2012.
- [5] ETSI. (2014) Network functions virtualisation (NFV); architectural framework. [Online]. Available: https://portal.etsi.org/Portals/0/TBpages/NFV/Docs/NFV_White_Paper3.pdf
- [6] E. Liotou, A. Marotta, L. Pomante, and K. Ramantas, "A middleware architecture for QoE provisioning in mobile networks," in *2017 IEEE 22nd CAMAD*, June 2017, pp. 1–5.
- [7] A. Virdis, G. Stea, and G. Nardini, "SimuLTE - a modular system-level simulator for LTE/LTE-A networks based on OMNeT++," in *2014 4th SIMULTECH*, Aug 2014, pp. 59–70.
- [8] Z. Bojthe, L. Meszaros, B. Seregi, R. Hornig, and A. Varga, *INET Framework for OMNeT++*, 2016.
- [9] D. Ciabrone, S. Tennina, D. Tsolkas, and L. Pomante, "A QoE performance evaluation framework for LTE networks," in *2018 IEEE WoWMoM - Workshop on User-centric service provisioning over 5G networks (USEPRO'18)*, Chania, Crete, Greece, Jun. 2018.
- [10] J. A. Bergstra and C. Middelburg, "ITU-T recommendation G.107: The E-Model, a computational model for use in transmission planning," 2003.
- [11] T. Falk and W. Chan, "Single ended method for objective speech quality assessment in narrowband telephony applications," *ITU-T*, p. 563, 2004.
- [12] ITU-T, "Recommendation G.1070: Opinion model for video-telephony applications," 2012.
- [13] D. Pal and V. Vanijja, "Effect of network qos on user qoe for a mobile video streaming service using h.265/vp9 codec," *Procedia Computer Science*, vol. 111, pp. 214 – 222, 2017, the 8th International Conference on Advances in Information Technology. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050917312292>