

Leveraging OTT and ISP cooperation to enhance end to end QoS by exchanging valuable resources

Michele Scarlato*

michele.scarlato@diee.unica.it

*DIEE, University of Cagliari, Italy

Jordi Ortiz†

jordi.ortiz@um.es

Cristian Perra*

cperra@ieee.org

Antonio Skarmeta†

skarmeta@um.es

†DIIC, University of Murcia, Spain

Abstract—Over The Top (OTT) provider detecting bad QoS for their users can not influence on the network by themselves. In the other hand, the Internet Service Providers (ISPs) along the path only see the effect in their own network. The users might agree on reducing a property (e.g., the bandwidth) on contract Service Level Agreements (SLA) with ISP to enhance other properties (e.g., Round Trip Time). This paper proposes a communication mechanism between OTT and its own ISP who could act as a relay with bordering ISPs chaining up to user premises, thus finally obtaining better QoS for the pair OTT/user reducing ISPs load in return. This paper proposes a REST interface for the specific case in which enforced RTT is to be exchanged by reducing available bandwidth as a counterpart, to be extended in future work.

Index Terms—Quality of Service (QoS), Service Level Agreement (SLA), Over-The-Top (OTT) service, Internet Service Provider (ISP).

I. INTRODUCTION

Two entities exchanging information remotely is defined as end-to-end connectivity. This is true if we omit the fact that in world wide communications this connectivity is delivered over a bunch of ISPs. In general and for short term communications, the entities have only control over the last mile connectivity parameters by means of contracts with strict SLAs with their ISPs.

If specific QoS is required for a certain service, only in special cases and mainly through third parties, long term SLAs can be signed, but for individuals and for short periods of times, this kind of service is not affordable. As a result, the deployment of services in the last mile such as CDNs has been lately widely employed. But for interactive services in which content is dynamic and/or calculated, the prior approach is not always feasible.

OTTs operate on top of the Internet and those relying in end-to-end communications between their customers and centralised servers suffer from the best effort approach of the Internet. Even if the OTT might be able to identify low QoS for a certain customer which probably is not a customer from its own ISP, the OTT can negotiate better SLAs with its own ISP and obtain better communications for the whole connection but that does not ensure that this precise customer obtains an increase in the QoS. The OTT requires the means to request for enhancements for a certain customer which in turn might require extending the request to multiple ISPs in the path in a delegated QoS request. In this paper we

define a Representational State Transfer (REST) application program interface (API). The APIs we defined use the HTTP requests POST and GET, in order to create and retrieve resources. RESTful API is an architectural style and approach to communications that is generally used in the development of web services. The stateless nature of the calls enable this approach to scale and accommodate load changes.

The customers on the other hand might agree on reducing some terms of their SLAs to augment others, e.g, reduce the download speed in order to reduce the RTT to a certain service, similarly to what some DSL operators offered with Annex M by which upload speed was incremented by means of sacrificing the download speed on user request.

How the changes in the network might be enforced is out of the scope of this paper but leveraging on SDN looks like a good starting point for such a system.

These dynamic agreements attract new business models in which OTTs would be able to enhance their Quality of Experience (QoE) by means of customer oriented QoS enhancement, users could experience an increase in the QoS for their favourite OTT by reducing unused parameters of their contracts and finally the ISPs, not only the one related to customer and OTT, but also the intermediate ones which could sell better paths to the edge ISPs. These last ones would be able to reduce the over-provisioning of their networks by means of these agreements with their customers, exchanging SLA capabilities in a profitable way for the ISP and with extra incoming in the case of the OTT by offering a more fine grained service enhancement.

The rest of this paper is organized as follows. Section II describes related works. In Section III, the proposed solution is described whereas in Section IV conclusions are drawn.

II. RELATED WORKS

Many solutions have been proposed in order to solve QoE and QoS related problems. Among them, there are solutions that make use of the Software Defined Networks (SDNs), which is a network paradigm where the programmability of the control plane can be exploited in order to monitor QoE parameters. Other solutions taken into consideration in our work have an economical approach.

The generic QoE management framework presented by Seppanen et al. [1] which is applicable to a broad range of systems can be useful for data acquisition and the monitoring

levels. In their work it is demonstrated an instantiation of the framework as a network access point management system for RTP-based video. The system is able to positively affect the perceived quality of the multimedia application considered, and also to reduce over-prioritization and optimize resource usage.

The QoE function proposed as a “QoE-service” for on-demand services or premium users based on SDN introduced by Liotou et al. [2] is able to provide a global resource view. Furthermore the view is combined with complementary QoE metrics to assure the desired performance for OTT applications by adopting traffic management mechanisms. Also they investigated a set of use cases that demonstrate its suitability and applicability to Long Term Evolution (LTE) networks.

An interesting management and orchestration architecture able to differentiate network services with quality level assurance and to enforce agreed SLA has been defined by Ongaro et al. [3]. The defined architecture exploits the use of SDN in conjunction with the OpenFlow protocol. They formulate the problem of enhancing QoS and QoE in terms of packet loss and delay, with the Integer Linear Programming (ILP). In the formulation the network constraints and the requirements of real-time applications have been taken into account. Also, once discovered the optimal solution, they evaluate the impact and benefits of the proposed schema using the Mininet network emulator.

Durner et al. [4], provide a study of the impact on dynamic QoS mechanisms and their realizations for OpenFlow-enabled SDN switches. Although SDN and, in particular, OpenFlow as one dominant realization claim to provide a standardized interface to control network traffic, their measurement results show a noticeable diversity for different OpenFlow switches.

III. PROPOSAL

In our proposal it is considered the general use case where both user and the OTTs perceive a lack of quality in the service. But the OTTs might not be able to influence, neither know, anything about the user’s ISP connection, even if the ISP is the same for both entities. For the OTTs, receiving information related to the connection of the users would be very useful to figure out what the problem is.

We propose a way to trigger the ISP to make an exchange of one parameter for another, for example enabling the exchange of bandwidth in excess in favor of a reduction of the RTT, in order to provide a better QoS to final users. In particular gaming, video streaming and video-conferences are applications that need to reduce the lag.

We are proposing REST APIs useful for communication among OTTs and ISPs, able to show to the OTTs some QoS parameters, and verify if they are below the desired levels, meanwhile other parameters are being wasted or unused.

There is no restriction or suggestion on how the ISPs could enforce a certain QoS level and restrain the exceeding parameters in exchange, but it is clear that the Software Defined Networking paradigm can play a key role in this kind of solutions, both because of the possibilities that it offers in

terms of applying flow based policies as well as the centralized control view of the network therefore being able to assess its real status.

In table I are described the calls to the methods that we consider useful in order to provide a better QoS to final users experiencing lag in the gaming use case.

The simplest case we considered is when OTT and user are using the same ISP. Whenever there are one or more ISPs in the middle, we have to consider the chaining. The chaining here makes reference to take advantage of the connection between border ISPs which implies a trust relation so that the OTT’s ISP relays the request if accepted to the next ISP likewise packets are forwarded based on routing algorithms. It is not foreseeable a direct connection between the OTT and any ISP which is not providing a direct service to the former.

For example, if the game company is Spanish and connects through Telefonica, and the user is connected through the same telecommunication company, the ISP will be the same.

Supposing that the ISP and the game company deployed the REST APIs that we are proposing, and the latter notices that one or more of his users are experiencing lag, the OTT will try to ask the ISP to reduce the RTT for a certain user, or a group of users.

The gaming OTT can perform a call to the method `POST /qos/rttreduction?network=IPv4/netmask&maxrtt=maxdesiredrtt` asking the ISP to reduce the RTT, passing as parameters the IP of the host or the network of hosts, and the maximum RTT desired.

The information regarding the QoS parameters of the user connection can be retrieved by the OTT calling the method `GET /qos/minmaxrtt?network=IPv4/netmask`, which is used to request to the ISPs the minimum and the maximum RTT for a particular host, or network of hosts. In order to have a fair exchange, reducing the lag in exchange of, for example, a reduction of the bandwidth, the OTT needs to know which is the amount of bandwidth available for one particular host, or a network of hosts. The game company, in this case, can call the method `GET /qos/availablebw?network=IPv4/netmask`.

If there is remaining bandwidth for a certain user, or a group of users, and it is possible to reduce its or their RTT, the OTT can propose to the ISP to exchange the available bandwidth with a reduction of the RTT, calling the method `POST /qos/exch-av-bw4rtt-red`.

In scenarios where the OTT and the User are connected to different ISPs, chaining concept as shown in Figure 1 must be applied. It is shown how the OTT is connected to its ISP, in turn connected to bordering ISPs. The former before reaching the users ISP has to cross one or more border ISPs in the path to offer End-to-End QoS. We assume in this scenario that data concerning the QoS perceived by the gaming OTT for certain User is measured in the OTT server and is provided to a QoS Monitoring service. This latter triggers if necessary a QoS Reduction via a QoS Reaction module, which in turn contacts the ISPs External Services API (ESA) which could be implemented as a standalone service proxying to the

TABLE I
A REST API FOR ISP COMMUNICATION IN QoS...

REST Call	Description
<i>POST /qos/rttreduction?network=IPv4/netmask&maxrtt=maxdesiredrtt</i>	With this POST the gaming OTT asks the ISP to reduce the Round Trip Time, passing the IP of the host or the network of hosts, and passing the maximum RTT desired.
<i>GET /qos/minmaxrtt?network = IPv4/netmask</i>	With this GET it is requested to the ISP the minimum and the maximum RTT for one particular host, or network of hosts.
<i>GET /qos/availablebw?network = IPv4/netmask</i>	With this GET it is requested to the ISP the bandwidth available for one particular host, or network of hosts.
<i>POST /qos/exch - av - bw4rtt - red</i>	With this POST the OTT wants to exchange the available bandwidth with a reduction of the RTT.

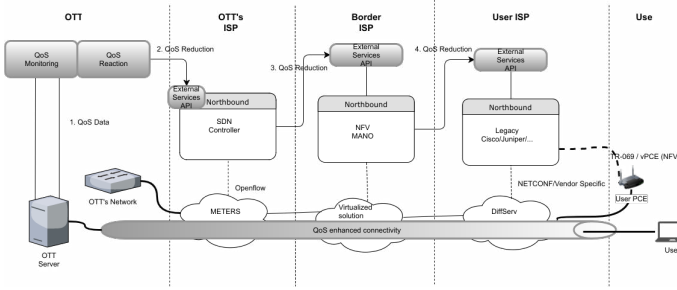


Fig. 1. Chaining requests.

network controlling system (let it be SDN or any other network management system) or integrated as part of the northbound interface of the controlling system itself. The ISPs network management system will check if it is possible to enhance the QoS in its domain and will forward a request to the next (if any) ISP in the path to the user. It can be assumed a certain degree of collaboration between bordering ISPs. Once the request arrives the Users ISP and if SLAs with the user in addition to network architecture allow it, the QoS enhancement could be exchanged for exceeding bandwidth, be billed, or any other prestablished agreement. The figure exemplifies some of the actual and probably future network architectures that are foreseeable to be employed by ISPs but does not try to be complete in that sense.

So, the gaming OTT will trigger a call to the POST where each ISP that is in the path is asked to reduce the RTT. Also the following steps have to be performed for each ISP: the GET requesting the minimum and the maximum RTT, the GET requesting the bandwidth available, and the POST where it is expressed the will to exchange the available bandwidth, whereas there is some, for a reduction of the RTT, whenever it is possible.

IV. CONCLUSIONS

This work proposes the basis and a first approach to a chained collaboration between OTTs and ISPs focused on the enhancement of QoS by exchanging and balancing connection parameter values, so that the OTTs provide with enhanced service to their customers by a reduced price or even free, the ISPs provide with better QoE to the customer while reducing

parts of their SLAs as a counterpart and the customer finally solving a problem with the service being consumed almost transparently.

The RESTful methods proposed are envisioned for the use case in which RTT is requested to be reduced by accepting a downgrade in the available bandwidth as a counterpart. We find these calls interesting for services like online videogames or videoconferencing systems, among others. The contrary might be also interesting for other use cases and many other exchanges will be also inspected as part of our future work.

The definition of an instantiation of this proposal and specifying in detail how it could be achieved by employing SDN is also part of our schedule. Finally, the biggest challenge is how to model a SLA with the client that would allow this kind of exchanges so that it can be enforced by ISPs, for that legal advice from researchers in legal areas will be pursued.

V. ACKNOWLEDGEMENTS

This work was partially supported by the European Commission through the Erasmus Mundus Action 2 Project TEAM Technologies for Information and Communication Technologies, the PhD scholarship from Region Sardinia (P.O.R. Sardegna, European Social Fund 2007- 2013 - Axis IV Human Resources, Line of Activity 1.3.1.) and by the H2020 EU project ANASTACIA project, Grant Agreement N 731558 as well as by the GN4-2 project under Grant Agreement No. 731122.

REFERENCES

- [1] J. Seppänen, M. Varela, and A. Sgora, "An autonomous QoE-driven network management framework," *Journal of Visual Communication and Image Representation*, vol. 25, no. 3, pp. 565–577, 2014.
- [2] E. Liotou, G. Tseliou, K. Samdanis, D. Tsolkas, F. Adelantado, and C. Verikoukis, "An sdn qoe-service for dynamically enhancing the performance of ott applications," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*. IEEE, 2015, pp. 1–2.
- [3] F. Ongaro, E. Cerqueira, L. Foschini, A. Corradi, and M. Gerla, "Enhancing the quality level support for real-time multimedia applications in software-defined networks," in *Computing, Networking and Communications (ICNC), 2015 International Conference on*. IEEE, 2015, pp. 505–509.
- [4] R. Durner, A. Blenk, and W. Kellerer, "Performance study of dynamic qos management for openflow-enabled sdn switches," in *Quality of Service (IWQoS), 2015 IEEE 23rd International Symposium on*. IEEE, 2015, pp. 177–182.