Performance evaluation in ns-3 of a video delivery framework for next generation cellular networks

Daniele Munaretto†*, Marco Zanforlin*, Michele Zorzi†*
† Department of Information Engineering, University of Padova, Italy
* Consorzio Ferrara Ricerche, Ferrara, Italy
E-mail: {munaretto, zanforli, zorzi}@dei.unipd.it

***Abstract*—Video mobile applications can be served via multiple delivery paths from the video server to the end user, thus delivering different video qualities. Throughput and delay highly depend on the video paths available in the core network and on the availability of wireless access technologies in the last hop. We consider a set of characteristics of the whole video delivery chain to univocally identify each available path and we develop a framework for the selection of the best video path in terms of throughput and packet delivery delay. We further extend the framework, implementing at the access points a tunable traffic shaping mechanism to decrease the delivery delay while maintaining the number of packets delivered. We evaluate our framework by means of simulation in ns-3, an event-based network simulator able to accurately model the Long Term Evolution (LTE) core and access networks.**

***Index Terms*—Video quality; LTE; WiFi; path selection.**

I. Introduction

The tremendous increase of mobile video traffic generated by the growing range of multimedia applications triggers the community to face new transport optimization challenges. From the mobile operator's point of view, one of the main challenges is the emergence of novel billing models which need to take into account the availability of heterogenous broadband wireless access technologies for the end users while targeting a significant reduction of the network resources and operational costs. The mobile operator is further asked by video consumers to meet some user satisfaction criteria, which is a complex challenge to be addressed due to the sophisticated requirements of the current and future emerging services, that go well beyond the traditional Quality of Service (QoS) provisioning of the past years. The deployment of wireless local area network (WLAN) hot spots to complement cellular access represents a key solution for the operators, since part of the mobile traffic that cannot be sustained by the base stations can be off-loaded to WiFi. Nevertheless, this also impacts the core network, which is required to meet greatly increasing capacity demands. This way, the whole video delivery chain is responsible for maintaining the video quality, and transport optimization solutions should be designed in an end-to-end (E2E) fashion, from the video sources to the video consumers. Existing solutions typically target the optimization of well defined and limited areas of interest, without considering the impacts of each of the proposed solutions on the other segments of the video delivery chain, potentially leading to some mismatching among them.

This work was supported by MEDIEVAL (MultiMEDia transport for mobIlE Video AppLications), a medium-scale focused EU research project of the 7th Framework Programme of the European Commission, grant agreement FP7-258053.

Our idea is to consider the availability of multiple E2E video paths, i.e., from a video source to the end user, and to select among them the path that best meets the user satisfaction in terms of two performance metrics: (i) throughput, defined as the total number of packets successfully delivered to the users in a video session, and (ii) packet delivery delay, defined as the E2E delivery time measured from the source to the end user. Thus, we design a framework that enables the selection of the video path and makes use of traffic shaping techniques at the access points to avoid network congestion due to sudden unbalanced heavy data traffic, e.g., before the traffic redirection is made effective. Our goal is to keep the throughput of mobile users as high as requested by quality demanding applications, e.g., guaranteed bit-rate (GBR) services, while reducing the delivery delays experienced by the users. Thus, we cover a set of applications ranging from real-time video streaming, with strict delay requirements, to video-on-demand (VoD), characterized by higher throughput and more relaxed delay constraints. The performance of our proposed mechanisms is investigated, in terms of throughput and packet delivery delay, for typical pedestrian mobile users which can receive their video packets from multiple video sources available in the network via either the WiFi or the LTE [1] access network. We evaluate the performance by means of simulation in ns-3 [2], an event-based network simulator implementing the LTE core and access networks.

The remainder of the paper is organized as follows. In Section II we review the prior work. We describe our reference scenario in Section III and our proposed framework in Section IV. The simulation setup and results are provided in Section V and we conclude the paper in Section VI.

II. Related Work

Mobile video services designed for next generation cellular networks, i.e., LTE [1] and LTE-A (LTE-Advanced) [3], have attracted the interest of the community. Mobile video consumers are equipped with sophisticated smartphones which can combine wireless access technologies independent of each other in order to realize the feeling of being always-online, no matter which access technology is available.

Prior solutions for video transport optimization over cellular networks cover different areas, where each of them is investigated and developed independently of the others. For instance, in [4] a cross-layer scheduler for video transmission is presented, focusing on the wireless access side only, without taking into account the network resources as a whole. Network operators target the efficient utilization of the overall network resources while providing high quality video transport, thus requiring novel designs and raising multidisciplinary topics in the community [5]. In a mobile network, a video application can be served through different video paths, taking into account both the core and the access networks, resulting in different quality perceived by the video consumer [6]. This is due to the response time of the core network, to the lossy nature of the wireless channel and to the available data rate for video delivery [7] on the wireless access side. A first combination of video transport optimization mechanisms considering the complete delivery chain over mobile networks was proposed in [8], where interworking traffic shaping mechanisms are conceptually discussed.

In contrast to the aforementioned related works, we extend the design of the path selection algorithm in [6] by taking into account new network metrics in order to cut the packet delivery delay while keeping, or even improving, the target data rate requested by the video service providers. We then implement our delivery framework in ns-3 to accurately evaluate the impact of the system on an LTE-compliant network, thus improving the first analytical evaluation in [6]. Finally, we make use of traffic shaping techniques at the access points to further reduce the delivery delays with a negligible impact on the throughput.

III. Architecture

The framework implemented in this work interacts with the mobile network architecture of Fig. 1. We consider the LTE cellular network from the access side to the Evolved Packet Core (EPC) network. In currently implemented solutions, mobile users can access the video contents through either the eNodeB or the WiFi hotspot, depending on the channel quality measured at the terminal for the two technologies. A mobile operator can deploy a set of Content Delivery Network (CDN) nodes, which act as interchangeable video sources, connected to the Packet Data Network (PDN) gateways (P-GWs), which give Internet access to the mobile device. Thus, the required video content can be potentially sent from alternative sources, each having different properties in terms of storage occupancy and distance from the access points, i.e., eNodeB and WiFi spot.

Our idea is to leverage on the LTE or WiFi availability for wireless access and on the multiple choices of video sources from where to download the content in the core network to best route the video packets to the end user in terms of throughput and packet delivery delay. Traffic engineering mechanisms implemented at the access points in order to decongest the queues of packets further impact the performance metrics and give a further degree of freedom to the mobile operator when tailoring the system performance to the characteristics (delay constraints, guaranteed bit-rate) of a target video service.

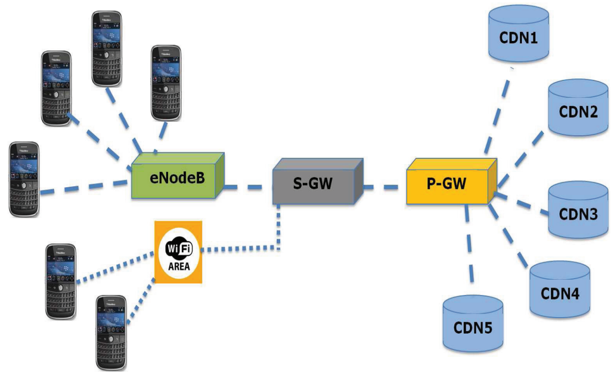


Fig. 1. Reference architecture: mobile users can access video contents from a set of video sources via LTE or WiFi.

A. LTE in ns-3

ns-3 is an open source discrete-event network simulator written in C++. The LTE module architecture [9], that was integrated starting from the stable release ns-3.14, is described in [10], while the current development of the third generation partnership project (3GPP) compliant framework can be downloaded from [2]. Two main components build the LTE mobile network in ns-3. The first is the LTE model which includes the LTE Radio Protocol stack. The second component is the EPC model, which includes core network interfaces, protocols and core network nodes. The interested reader can find further details in [2].

IV. Proposed Algorithm

The goal of our work is to investigate the impact of a video delivery framework, working on the whole delivery chain, on throughput and packet delivery delay for a set of video services. This way we provide the guidelines for fine-tuning the network metrics involved and for steering the video path to best meet the requirements of the end users and of the service providers. The search for a video path that maximizes the measure of the proximity to an ideal tuple, e.g., formed by the most desirable value taken by each network metric describing the video path, was initially studied in [6]. The aim to maximize a utility function on the "normalized" M-tuples representing all the possible paths (where each of the M values in a tuple refers to a given network metric) is now further investigated from a deployment point of view. To do so, we investigate the feasibility of the algorithm in an LTE-compliant network implemented in ns-3, we add further network metrics to increase the accuracy of such algorithm and we introduce traffic engineering mechanisms to support the video path selection in practical scenarios.

A. Path Selection

In [6] two criteria for the selection of the video paths were investigated, reflecting the operator's and the user's points of view. Due to the better performance achieved by the *max-sum* criterion, we select for this work the utility function maximizing the sum of the proximity values associated to the metrics in the M-tuples. The interested reader can find the details of the *max-sum* pseudo-algorithm in [6], for a generic number of metrics assigned in the tuples. In this framework we consider the following network metrics: the storage occupancy of a CDN video server, the routing distance from the server to the points of access, the effective data rate on the link between the video source and the P-GW (taking into account concurrent sessions) and the channel quality of the wireless link. The storage occupancy of the CDNs is computed as the fraction of the memory in use. The routing distance is expressed as the number of hops between the CDN cache and the end user. The data rate on the link between the CDN video source and the P-GW is also taken into account, since once a CDN is selected for transmission, the available bit-rate on the link should count for the corresponding bandwidth consumption due to concurrent sessions, previously established by other nodes, on the same link. The values taken by these core network metrics can be communicated in the network via ALTO protocol extensions enabling joint transmission of multiple metric values as proposed in [11]. Then, we consider the Signal-to-Noise Ratio (SNR) of both cellular and WiFi access as a quality measure of the wireless link, observed by the mobile terminal from the wireless interface. We use the same mapping as in [6] of the network metrics to the application performance metrics, i.e., the channel capacity and the response time since the effective bit-rate on the link between CDNs and P-GW has only impact on the selection mechanism of the video path. In the rest of the paper this algorithm will be referred to as Path Selection algorithm (PS) and will be investigated in Sec. V both as a stand-alone mechanism and jointly with the Packet Dropping with Threshold mechanism (PDT), described in the following section.

B. *Packet Dropping with Threshold*

Existing traffic engineering mechanisms are enabled in the mobile network when congestion occurs. Based on the video specifications, operators might decide to reduce the data rate from the source, in the core network nodes or at the access points. The selection of the video path for a set of end users might cause the temporary congestion of some preferred links/nodes, leading to a degradation of the network performance and of the quality perceived by the user. We thus shape the video traffic at both the eNodeB and the WiFi spot by dropping the video packets before being queued in the buffers of the access points. We notice that the congestion lasts until a new available video path is selected. Since the procedure could involve several links and nodes throughout the mobile network, it is assumed to be solved after a considerable number of packet transmissions. The mechanism of packet dropping is implemented as follows. When the end-to-end delay of the packet received by a given end user exceeds a fixed threshold, a mechanism in the corresponding access point is triggered such that the incoming packets from the core network and meant for any other node are dropped before being queued in the buffer(s) of the access points. This makes room for packets to be sent to the node observing high delivery delays. Once the measured packet delay falls below the threshold, the packet dropping procedure is disabled (otherwise a timer is triggered when the congestion occurs and expires after a fixed time interval).

V. Simulation Results

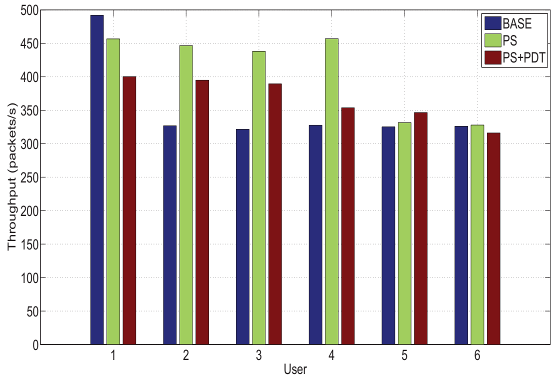
A. *Simulation Setup*

We implement the PS and PDT algorithms in the LTE module in ns-3. In our reference scenario we deploy 5 CDN sources in the core network (this number of CDN nodes was found to be near-optimal in [6]), connected to the P-GW, and we deploy 6 mobile users with different mobility patterns in the wireless access network. The LTE access network is implemented as follows. We consider an LTE micro-cell with bandwidth ranging from 5 to 20 MHz (i.e., number of resource blocks between 25 and 100) and a coverage radius of up to 5 km. The transmission power of the eNodeB is set to 30 dBm, the noise figure to 5 dB, and we use the Friis channel propagation model. The buffers dedicated to each user to be served by the eNodeB have limited size set to 2^{21} Bytes. We implement a WiFi 802.11a spot, with bandwidth set to 20 MHz, at a distance of 3 Km from the eNodeB and with coverage radius of up to 200 m. The transmission power of the access point is set to 16 dBm, the noise figure to 7 dB, and the channel propagation model is the log-distance model. The common buffer to serve all the users attached to the access point is limited to the size of 2^{20} Bytes. We implement in the EPC module the CDN caches such that each CDN has its own storage capacity and the link between each CDN and P-GW is set to the same channel rate, but different propagation delays due to the geographic locations. The video sources generate video packets at a rate of 500 KB/s, with packet size set to 1024 Bytes, while the delay on each CDN-P-GW link is randomly chosen in the interval $[1, 500]$ ms. We let the delays of the links change, with values picked from the same time interval, every 2 s.

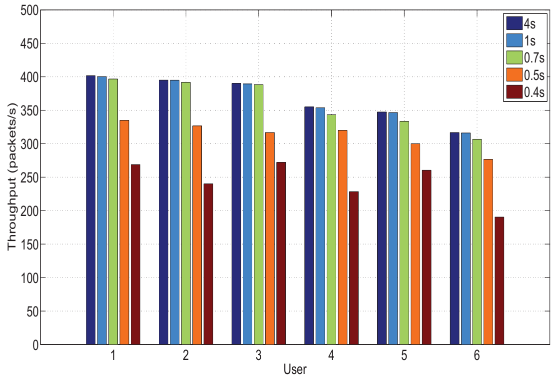
The users' mobility pattern is implemented as follows. Users move following a way point mobility model, making sure that they cross the WiFi area at least once within each simulation run and they are always in the LTE coverage area. The average speed of the users (urban area) is set to 2 m/s. We notice that, once the user is in the range of the WiFi hot spot, an additional access technology is available, thus increasing the chances of maintaining as good as possible the performance of the ongoing video session. Moreover, from an operator's point of view, this leads to an additional capacity to redistribute to the users that are not covered by the WiFi spot. As a final remark, while PDT runs instantaneously, i.e., as soon as the measure of the packet delay exceeds the threshold, the PS algorithm operates every 2 s, a value that takes into account the signaling propagation and collection of application-layer traffic optimization (ALTO) messages containing the network metrics, i.e., inputs of the PS [6].

B. *Discussion*

In this section we discuss the performance of the aforementioned algorithms in terms of throughput and packet delivery delay measured at the end user compared to a baseline solution, named "Base", implemented in the latest release of ns-3 as an LTE-compliant video transport mechanism developed according to the 3GPP specifications. There, a mobile user is connected to either the eNodeB or the WiFi spot and receives

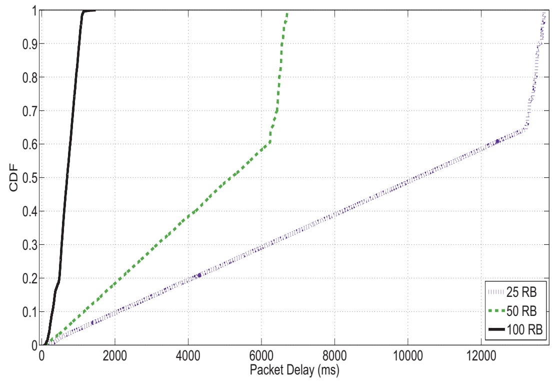


(a) Throughput measured at the nodes for Base, PS and PS+PDT (from left to right in each group), with threshold set to 1s.

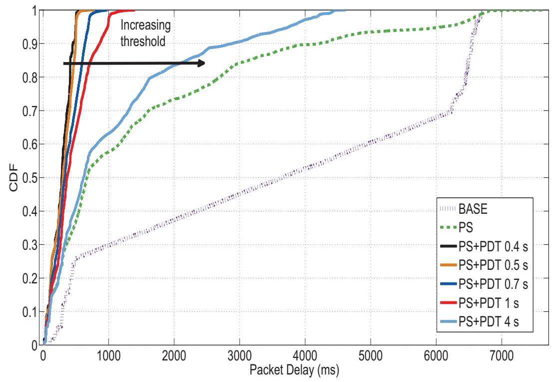


(b) Throughput measured at the nodes for PS+PDT, with different thresholds (descending order from left to right in each group).

Fig. 2. Impact of the Base, PS and PS+PDT algorithms on the throughput for the six mobile users.



(a) Cumulative distribution function of the average packet delays measured at the nodes for number of RBs set to 25, 50 and 100, baseline case.



(b) Cumulative distribution function of the average packet delays measured at the nodes for number of RBs set to 50.

Fig. 3. Impact of the Base, PS and PS+PDT algorithms on the packet delay of the mobile users.

the video packets uniquely from the first video source selected at the time of the session initialization. In our simulations we take into account the impact of mobility and load of the system to evaluate the system in terms of throughput and packet delivery delay, key metrics to assess the performance of real-time video streaming applications. In the plots, we compare Base, PS and the combination of PS and PDT, namely "PS+PDT", algorithms. This way, we provide a set of network management solutions, which offer different trade-offs in terms of throughput and delay. We argue that an increasing performance of the network as we enable in sequence the PS and the PDT traffic engineering mechanisms comes at the cost of an increasing computational complexity of the system. Depending on the service and users' requirements, the mobile operator might deploy only the PS algorithm instead of the combination of PS and PDT, when a fine-tuning of the parameters involved in the framework can already meet the video service requirements.

In Fig. 2, we compare the throughput measured at the six mobile users when using the Base, PS and the PS+PDT algorithms. In Fig. 2(a), we plot the average throughput of the mobile users setting the threshold of the PDT to 1 s. To make the algorithms comparable, in the baseline case we let 5 users always connect to the eNodeB (and to the same CDN by definition of baseline algorithm) and 1 user to the WiFi access point. This way, the eNodeB is congested since the users cannot modify their initial choice of the best video path (otherwise the video session would be dropped) nor benefit from the packet dropping at the access points. The results show that PS is beneficial in terms of throughput compared to the baseline case, since the nodes can switch from one access point to the other, improving the average quality of the channel, and as a result the packet delivery rate. The combination of PS and PDT is slightly worse than PS in terms of throughput, as expected, but still acceptable for GBR applications, when setting, for instance, the target rate to $\sim$ 330 KB/s. However, as plotted in Fig. 2(b), setting a higher threshold corresponds to an increase of the throughput to values similar to the PS case, even though PDT drops packets from the video streams. When congestion occurs, PDT drops the packets before being queued in the buffers of the access points. The congested nodes, after a short time interval, benefit from such packets dropped in the buffers of the eNodeB or in the common buffer when considering the WiFi spot. Once the congestion is solved, more packets can flow again through the access points before the end of the video session, set to 30 s. When the threshold is set around 1 s, the percentage of dropped packets still makes it possible to increase the number of packets being delivered in time to the user before the end of the video session. This turns into a throughput gain for such high thresholds, at the cost of a slight increase of the packet delivery delay. Setting the threshold below 1 s, the percentage of dropped packets outweighs the ratio of packets "gained" by the system once the congestion is over, and the net result is an overall performance degradation.

Before discussing the impact of the algorithms on the packet delivery delay, we present in Fig. 3(a) the cumulative distribution function (CDF) of the packet delivery delays at the end users when the number of resource blocks (RBs) in use at the eNodeB by the baseline solution is equal to 25, 50 and 100. Hence, we analyze at which bandwidth the eNodeB is congested in the reference scenario. With 100 RBs the network delivers the packets avoiding congestions, while with 50 RBs the eNodeB is congested and half of the packets are delivered with delays above 6 s (even more with 25 RBs). For the sake of comparison, we target a scenario where the access points, at some point in time during the simulation, are fairly congested, thus we set the number of RBs to 50 in the simulation runs. The WiFi access point, at the bandwidth of 20 MHz, with six users crossing its coverage area, is rarely congested, giving the operator a chance to offload the traffic when congested or, from the user's point of view, to re-direct the video path to the next best available technology. In Fig. 3(b) we compare Base, PS and PS+PDT by showing the CDF of the packet delivery delay experienced by the users. As expected, in the baseline case the users experience the worst delays due to the pre-selected video path which is kept during the whole video session, thus congestions or slow links between the CDN selected and the P-GW affect the performance. The flexibility offered by PS makes it possible to deliver half of the total number of video packets in the system below 1 s of delivery delay. Further improvements are brought when PDT is enabled in the access points, leading to remarkable gains in terms of delivery delay, but at the cost of detrimental performance in terms of throughput for stringent thresholds, such as 0.4 and 0.5 s (see Fig. 2).

We compare in Fig. 4 the average throughput and packet delivery delay experienced by the users when using the different algorithms proposed. It can be observed that PS alone allows to significantly cut the delays with respect to the baseline case without affecting the throughput, which instead benefits from the video path redirections. PDT jointly with PS further cuts the delays in the delivery system while keeping the target delivery ratio, unless the threshold is set at very low values, thus making the network unstable. We remark that PDT alone is out of the scope of this paper since our main goal was to evaluate the PS algorithm in ns-3 to assess the impact of the algorithm on an LTE-compliant network simulator. Moreover, the target was to cut the packet delivery delays without significantly affecting the throughput of the system to meet the highly-demanding requirements of the latest emerging mobile services. We recall that it is possible for the operator to fine-tune the weights of the metrics taken into account in the PS algorithm, to adapt the performance of throughput and delivery delay to the dynamics of the network and to the targeted performance and available network resources.

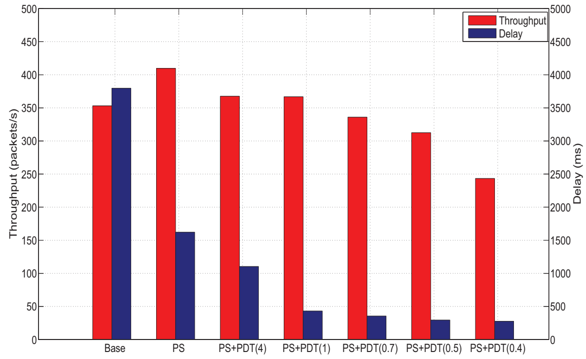


Fig. 4. Average throughput and delay comparison. In brackets the threshold value for PS+PDT.

VI. Conclusions

In this work we implemented and evaluated in ns-3 a delivery system to improve the video transport over LTE cellular networks with the goal of keeping a target system throughput while decreasing the packet delivery delay experienced by the mobile users. By selecting the best video path available for a set of network metrics, PS achieves remarkable gains in terms of delivery delay and slightly enhances the overall throughput of the system. When PDT is enabled jointly with PS, the delivery delay can be further decreased but at the cost of lowering the throughput and of additional computational complexity. At the time of writing we are testing the framework on a testbed reproducing the LTE mobile network. Hence, the mapping of the QoE-related performance metrics under study will be evaluated to eventually assess the impact of these mechanisms on the perceived quality.

References

[1] "3GPP Specification Detail: LTE." [Online]. Available: http://www.3gpp.org/LTE
[2] "ns-3 Network Simulator." [Online]. Available: http://www.nsnam.org
[3] "3GPP Specification Detail." [Online]. Available: http://www.3gpp.org
[4] W. Kumwilaisak, Y. T. Hou, Q. Zhang, W. Zhu, J. Kuo, and Y.-K. Zhang, "A cross-layer quality-of-service mapping architecture for video delivery in wireless networks," *IEEE JSAC*, vol. 21, pp. 1685 – 1698, Dec. 2003.
[5] Q. Zhang, W. Zhu, and Y. Q. Zhang, "End-to-End QoS for Video Delivery Over Wireless Internet," *Proceedings of the IEEE*, vol. 93, pp. 123 – 134, Jan. 2005.
[6] D. Munaretto, T. Melia, S. Randriamasy, and M. Zorzi, "Online path selection for video delivery over cellular networks," in *Proc. of QoEMC, IEEE Globecom 2012*, December 2012.
[7] S. Kahn, S. Duhovnikov, E. Steinbach, and W. Kellerer, "MOS-based multiuser multiapplication cross-layer optimization for mobile multimedia communication," *ACM Journal on Advances in Multimedia*, vol. 2007, pp. 1 – 11, Jan. 2007.
[8] B. Fu, D. Munaretto, T. Melia, B. Sayadi, and W. Kellerer, "Analyzing the Combination of Different Approaches for Video Transport Optimization for Next Generation Cellular Networks," *IEEE Network Magazine, special issue on video over mobile networks*, March 2013.
[9] "ns-3 LTE module." [Online]. Available: http://www.nsnam.org/docs/release/3.14/models/html/lte.html
[10] G. Piro, N.Baldo, and M. Miozzo, "An LTE Module for the ns-3 Network Simulator," in *Proc. of ICST SIMUTOOLS*, March 2011.
[11] "Multi-Cost ALTO." [Online]. Available: http://tools.ietf.org/id/draft-randriamasy-alto-multi-cost-05.txt