

Implementation of Multi-layer techniques using FEDERICA, PASITO and OneLab network infrastructures

V. López, J.L. Añamuro, V. Moreno,
J.E. López de Vergara and J. Aracil

High Performance Computing and Networking group
Universidad Autónoma de Madrid
Madrid, Spain.
E-mail: victor.lopez@uam.es

C. García and J.P. Fernández-Palacios

Telefónica I+D
Madrid, Spain
E-mail: jpfpg@tid.es

M. Izal

Universidad Pública de Navarra
Pamplona, Spain.
E-mail: mikel.izal@unavarra.es

Abstract—This paper describes an implementation of multi-layer techniques using the network infrastructure provided by FEDERICA, PASITO and OneLab projects. FEDERICA project provides a network infrastructure, based on virtualization capabilities in both network and computing resources, which creates custom-made virtual environments. PASITO is a layer-2 network that connects universities and research centers in Spain. OneLab measurements tools allow carrying out high-accuracy active network measurements. Thanks to FEDERICA and PASITO, we have a multi-layer architecture where the traffic is routed based on the measurements of OneLab equipment. To carry out this experiment, we have developed a Multi-layer Traffic Engineering manager and an implementation of the Path Computation Element Protocol to solve the lack of a control plane in IP oriented networks. This work shows the feasibility of multi-layer techniques as a convenient solution for network operators and it validates our Path Computation Element implementation.

Index Terms—Multilayer traffic engineering; Path Computation Element; Quality of Service.

I. INTRODUCTION

In recent years, the explosion of broadband connections has imposed an unprecedented traffic growth in telecommunication networks with very high cumulative annual growth rates. An example of this huge traffic growth is the forecast from Cisco [1] that predicts an annual IP traffic over 700 exabytes in 2014, four times higher than in 2009.

Current operator network deployments are mostly based on IP/MPLS routers, either on their own or supported by an optical switching network (WDM, OTN, etc.). IP and MPLS layers take advantage of statistical multiplexing of Internet traffic profile. However, new applications have lately appeared demanding higher capacity and changing the traffic profile. For instance, a file transfer does not have severe network requirements, while Video on Demand (VoD) streaming requires a minimum rate, delay and jitter [2].

In this new scenario, gain by statistical multiplexing is modified, specially in the network core. Therefore, the utilization of IP resources must be carefully organized. Multi-layer techniques try to use the resources of lower layer to

reduce congestion at the IP layer and to provide an efficient configuration of the IP resources. Previous work has shown that multi-layer techniques are feasible [3] from the technical point of view and they enable optimization of resources utilization in current IP backbone topologies [4], [5], [6].

The primary goal of performing multilayer coordination is cost reduction. By including an element applying multi-layer optimization, the operator can avoid inefficiency and reduce CAPEX. Transport and IP/MPLS transit capacity is required to interconnect border routers. However, their deployment can be jointly planned, leading to cost reduction. Whenever there is no sufficient gain from applying statistical multiplexing, the use of expensive IP equipment can be reduced. Some results of potential savings in the required investment for IP networks are shown in [3], [7], [8]. These studies demonstrate that significant CAPEX savings can be obtained by an appropriate combination of the resources in multiple layers. In previous work, authors in [3] developed a multi-layer prototype with three IP routers and three optical cross-connects to show the feasibility of multi-layer techniques. In this work, we carry out a multi-layer experiment involving a larger testbed thanks to FEDERICA and PASITO infrastructure with accurate timestamping using OneLab equipment. Moreover, we have developed an implementation of Path Computation Element (PCE) [9], which is validated with a second experiment. The PCE can help to integrate networks without a control with control plane oriented networks.

The remainder of this paper is organized as follows. Section II describes the multi-layer architecture and provides detailed information regarding the infrastructure of each project (FEDERICA, PASITO and OneLab), the Multi-layer Traffic Engineering manager and the PCE architecture. In Section III, the feasibility of multi-layer techniques is demonstrated on an experimental testbed with a video transmission. Moreover, we validate our Path Computation Element Protocol (PCEP). Finally, Section IV concludes the paper.

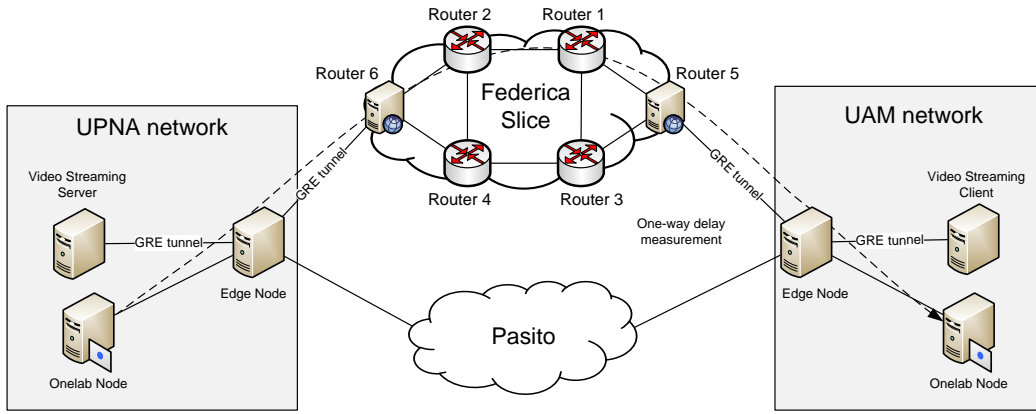


Fig. 1. Multi-layer network architecture used in the experiments

II. ARCHITECTURE OVERVIEW

Next Generation Transport Networks (NGTN) are comprised by a control, a management and a data plane [10]. The data plane is used for the transmission of information packets. Data plane is composed by the IP/MPLS or optical links where the user data is sent. The management plane deals with global operations, including accounting, security evaluation, monitoring reports, etc. The control plane is in charge of decentralized management issues such as the exchange of routing information, link state monitoring and the set up and tear down of connections. Additionally, such control plane manages the Service Level Agreements (SLA) and monitors the Quality of Service (QoS) offered to the connections.

The test-bed scenario is a distributed network communicating Universidad Autónoma de Madrid (UAM) and Universidad Pública de Navarra (UPNA) premises. The architecture uses resources from three facilities: FEDERICA, PASITO, and OneLab. The objectives of this work are to assess multi-layer techniques in realistic networks and to validate our implementation of the Path Computation Element. This section explains each part involved in the experiment: the multi-layer network, the monitoring probes, the coordination of all elements in the test-bed, the PCE architecture and its integration with control plane enabled networks.

A. Multi-layer network architecture

The multi-layer architecture for this experiment has a layer-3 (FEDERICA slice) and a layer-2 (PASITO) network (Fig. 1).

FEDERICA project [11] provides a network infrastructure that allows the creation of virtual scenarios to carry out networking experiments. As each scenario is isolated, multiple virtual scenarios can run at the same time in the infrastructure. These virtual scenarios are called “virtual slices”. A FEDERICA user can request a virtual slice with multiple nodes and each node can run a different operating system. The requested slice for our experiment is composed by six nodes: two Linux and four Juniper routers (Fig. 1). The Linux nodes are the edge routers of the slice (routers 5 and 6) and the Juniper routers are the core nodes (routers 1 to 4).

PASITO is a research network funded by Spanish Ministry of Industry, Tourism and Trade [12]. PASITO infrastructure provides a layer 2 network joining universities and research centers across Spain, but it is not the main network access for these centers. In fact PASITO network is isolated from Internet providing a single VLAN between associated research centers for research purposes. PASITO provides a layer-2 connection between edge nodes at UAM and UPNA premises (Fig. 1).

The coordination between both architectures must be done by the control or management plane. Our infrastructures do not support a control plane solution (like ASON or GMPLS). We have created a Multilayer Traffic Engineering (MTE) manager, which is described in section II-C, to coordinate both layers.

B. End-to-end Quality of Service monitoring

The available infrastructure from the OneLab project [13] allows high-accuracy network measurements. This capability is achieved thanks to an advanced network monitoring equipment, that includes the ARGOS network monitoring card developed at UAM. The ARGOS card is composed of two PCBs: a NetFPGA board and a sister card [14]. The NetFPGA is in charge of receiving and timestamping the incoming packets with the GPS information provided by the sister card.

Software-based network monitoring systems can use GPS information for node synchronization purposes. The timestamping task can be done at driver level inside the receiving machine. ARGOS card achieves a higher level of accuracy, because software-based timestamping not only introduces a measurement overhead (IO and networking stack delays) but also this overhead is highly variable due to the CPU load. ARGOS offers, roughly, a 10ns end-to-end accuracy thanks to a low-level timestamping system. Each packet is marked with the time information at the moment of its departure or arrival at the ARGOS card. Therefore, this timestamp value does not depend on the machine CPU consumption.

Two cards are required to measure the One-Way Delay (OWD). One ARGOS monitoring card sends UDP packet trains from a source to a destination node, as shown in Fig. 2. The ARGOS source card marks the packets with the departure time information in the UDP payload. When the packet reaches

the destination, the card includes the timestamp in an ARGOS header (Fig. 2). This timestamp is highly accurate since the marking is done when the packet is received. Moreover, ARGOS card is using the GPS information. The packet is sent to the host like a standard NIC. The host can compute the OWD with the timestamp mark at the source and destination nodes in the packet.

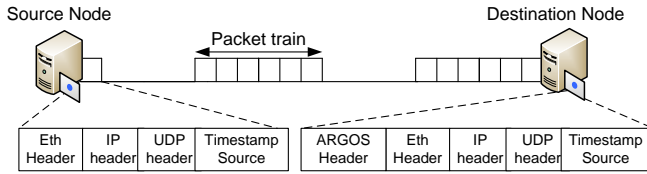


Fig. 2. ARGOS card train transmission and packet marking

C. Coordination between FEDERICA and PASITO networks

The coordination between the multiple layers in the network is done by the Multilayer Traffic Engineering (MTE) manager. The MTE manager is in charge of deciding where to send the traffic based on QoS information from the network. Layer 2 connections reduce the delay and traffic sent to the IP routers. On the other hand, the IP layer improves bandwidth utilization thanks to the statistical multiplexing. Flows to the same destination are merged, thus improving the efficient utilization of the resources. The MTE manager implements the following strategy. When the OWD in the IP layer is lower than a given threshold, the MTE manager sends the traffic using the FEDERICA network to reduce resource utilization, thus taking advantage of statistical multiplexing gains. When the IP layer is carrying a lot of traffic and the QoS constraints for the service are not fulfilled, the traffic is forwarded through PASITO network to provide the required QoS for the service.

An MTE manager is an integration element with the following generic requirements:

- 1) Access to monitoring tools of the IP/MPLS layer traffic.
- 2) Awareness of the routing information, as part of the control plane functions.
- 3) Implement multi-layer routing and restoration algorithms.
- 4) The configuration of the data plane and the control plane of the IP/MPLS network has to be made available in order to make a proper configuration of the IP/MPLS layer upon virtual topology changes.

Such element is required for a multi-layer architecture with a management and/or control plane. The MTE manager is an element performing integration of the management and control planes using QoS information. The obvious management protocol for monitoring and partially configuring IP networks is the Simple Network Management Protocol (SNMP) [15]. However, there are other scenarios with a control plane or monitoring probes that can provide QoS information to the MTE manager. Fig. 3 depicts the architecture of our MTE manager. Based on the previous list of requirements, we enumerate how our MTE manager fulfills them:

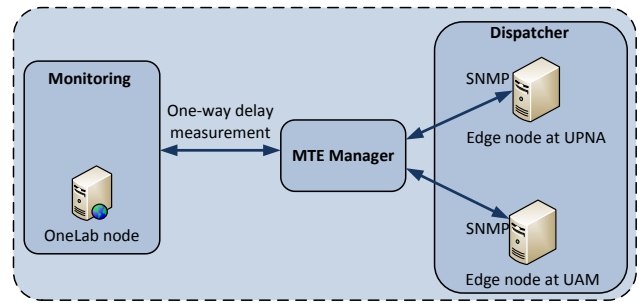


Fig. 3. Multi-layer Traffic Engineering manager

- 1) The MTE manager gets the monitoring information from a database that is feed by ARGOS cards with the OWD. ARGOS cards are used to continuously monitor FEDERICA slice (Fig 1). OneLab node at UAM premises fills in a database with the OWD in FEDERICA network from UPNA to UAM.
- 2) The MTE manager is aware of the current routing information at the edge nodes using SNMP.
- 3) The implemented algorithm routes the traffic using FEDERICA network, as long as the QoS requirements are fulfilled. When the delay is increased in the network, the manager sends the traffic using PASITO from UAM to UPNA premises.
- 4) The configuration of the data plane is done via SNMP modifying the routes in the edge nodes.

D. Path Computation Element architecture

The Path Computation Element (PCE) is “an entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints” [16]. Path Computation Element Protocol (PCEP) follows a request/response scheme as depicted in Fig. 5.

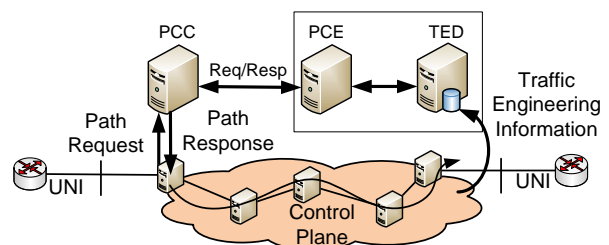


Fig. 5. Path Computation Element architecture integrated with a control plane

The Path Computation Client (PCC) requests routes to the Path Computation Element, who replies with the route. The PCE can be located as part of the management plane, where the network management system (NMS) requests paths to the PCE. If there is a control plane in the network, it asks for routes to the PCE. Once the PCE replies to the request, the NMS or the control plane can configure the routing elements. The PCE requires a traffic engineering database (TED) with the state information of the network. The TED can be filled

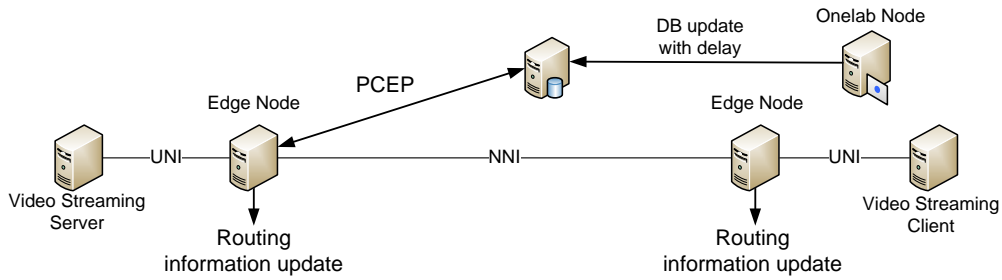


Fig. 4. Integration of FEDERICA-PASITO network with control plane networks

in with information from the NMS or the control plane. It is possible to monitor the status of the network via SNMP protocol or using monitoring probes (i.e. Argos card). When the control plane is used, the TED is filled in with the OSPF flooding information [17].

There are different PCE configurations for multi-layer networks defined in [18]. Basically, there are centralized and distributed models, where the routing computation is done in a centralized server or in multiple PCEs in the network. Moreover, there are single layer cooperating PCEs, where each PCE is in charge of a single layer, and multi-layer PCEs, which compute the route with the information of the whole network.

For our experiment, we have used a single PCE with a complete view of the multi-layer architecture. A more detailed explanation about the Path Computation Architecture can be found in [17].

E. Integration of multi-layer mechanisms in control-plane enabled networks

Neither FEDERICA nor PASITO has a GMPLS control plane enabled. However, in the last years a big effort has been put through the development of ASON and GMPLS. In networks with control plane, the MTE manager must have the availability of a User to Network Interface (UNI) in the boundaries between the IP/MPLS routers and the layer-2 network.

A possible integration of FEDERICA-PASITO network with control plane networks can be done using the Path Computation Element as shown in Fig. 4. The UNI interface receives the incoming request and it sends such request to the PCE. The PCE provides a response with the network configuration. The Edge Node process the path computation response to configure accordingly the routing tables via SNMP. The route information is sent via Network-to-Network Interface (NNI) to the edge node at UPNA, which updates the routing configuration in its node.

III. TESTBED SETUP AND EVALUATION

Fig. 1 shows the topology of the multi-layer network that was deployed with the OneLab elements and the video server and client in each site. GRE tunnels are used to connect FEDERICA slice with each network at UPNA and UAM. A GRE tunnel is created between UAM and UPNA edge nodes to connect them through PASITO network. We have carried out

two experiments using this test-bed: (A) assessment of multi-layer traffic engineering techniques and (B) validation of Path Computation Element Protocol.

A. Assessment of multi-layer traffic engineering

The first experiment consists on the video streaming from UPNA to UAM using FEDERICA and PASITO networks. The video server at UPNA transmits the traffic using FEDERICA slice, while the delay is lower than a given QoS threshold. According to [2], the delay for a streaming class service must lower than 500ms. The MTE manager is configured to send the video stream using FEDERICA, while the delay is lower than 500ms. To modify the delay, `netem` tool is used.

Fig. 6 illustrates the OWD from UPNA to UAM monitored by the ARGOS probes in this experiment. ARGOS cards are continuously transmitting a packet train and the destination OneLab node is measuring the average delay in the latest train. This average delay is inserted in the TED, so the MTE manager can decide where to send the traffic. If the OWD increases during the transmission, the MTE manager decides to send the traffic using PASITO infrastructure. Fig. 6 shows that there is a period where the delay is over the 500ms threshold.

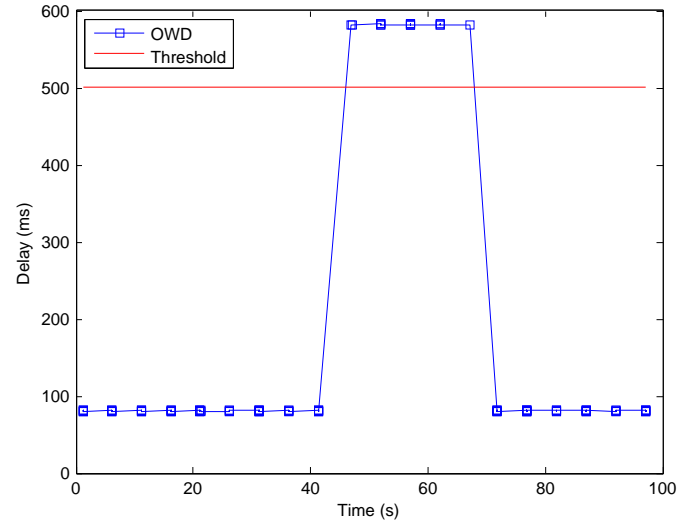


Fig. 6. End-to-end delay from ARGOS cards at UAM and UPNA

Fig. 7 illustrates the traffic received at UAM edge node. According to the figure, it is clearly shown that the routes

are switched when the OWD exceeds the threshold and the traffic is received not from FEDERICA interconnection, but from PASITO network.

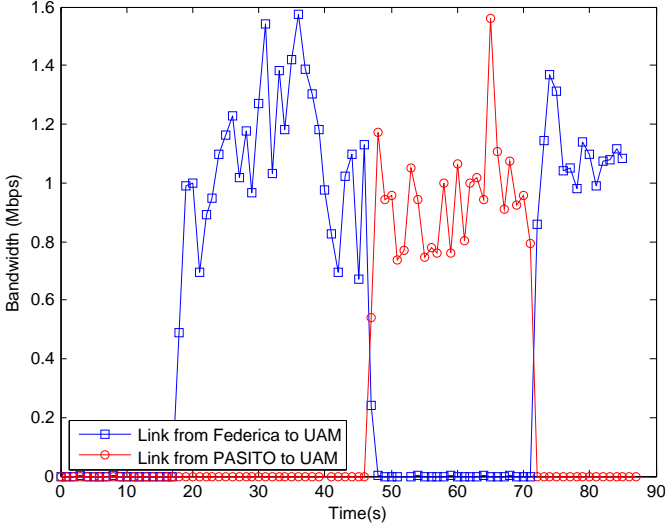


Fig. 7. Bandwidth used in FEDERICA and PASITO connection

Using this single threshold scheme, the routes can oscillate. To avoid such problem, a second threshold can be included. The second threshold must be lower than the first threshold (e.g. 400ms). With this second threshold, the traffic is reverted to FEDERICA when the delay is lower than 400ms.

B. Validation of the Path Computation Element Protocol

The second experiment validates the PCEP implementation with two path requests first without and then with congestion at FEDERICA. The PCEP is a request/response based protocol and it operates over the transport control protocol (TCP) [9]. There are seven possible messages namely Open, Keepalive, Request, Response, Notify, Error and Close [9]. The protocol follows the next steps:

- 1) **Session creation.** There is a establishment of a TCP connection (3-way handshake) between the PCC and the PCE using a TCP registered port (4189). Later there is the establishment of the PCEP protocol with a two-way PCEP Open messages. Using these messages, the configuration parameters are exchanged. To accept the PCEP connection Keepalive messages are sent.
- 2) **Session Keepalive.** Keepalive messages are sent to detect if the peers are available for use. A Keepalive message is transmitted every *Keepalive_timer*. There is a second timer namely *Dead_timer* which equals to 4 times the value of the *Keepalive_timer*. If there is no Keepalive message in *Dead_timer*, the TCP connection is closed. There are two operating modes for the PCEP: intermittent and permanent. In the intermittent mode, every time there is a request in the PCC, the PCEP connection is open and close. When the permanent mode is used, the PCEP connection is open and never closed. To maintain the connection, Keepalive messages are

sent every Keepalive timer. Only in the permanent mode, there is the session keepalive.

- 3) **Path Computation Process.** The *Request* message is sent from the PCC to the PCE, when a request arrives to the PCC. The PCE sends a *Response* message with the path computed based on the conditions included in the request message.
- 4) **Session close.** The PCE *Close* message is transmitted to close the PCEP session and then the TCP session is terminated.

Once PCE sessions are detailed, let us explain the second experiment. The TED is filled in with the OWD information as depicted in Fig. 4. Two path computation requests are done to the PCE. In the first request, the OWD at FEDERICA is lower than 500ms, while, in the second request, it is greater than 500ms. Consequently, the first computed route replies with IPs at the FEDERICA slice. The second request finds that the delay through FEDERICA is higher than the video QoS requirements so the PCE replies with the IPs of the PASITO nodes.

Time (s)	PCC (130.206.162.XX)	PCE (150.244.56.XX)
0.0000		OPEN MESSAGE
0.0210		OPEN MESSAGE
0.0211		KEEPALIVE MESSAGE
0.0416		KEEPALIVE MESSAGE
0.0417		REQUEST MESSAGE
0.0715		REPLY MESSAGE
0.0716		CLOSE MESSAGE
4.7703		OPEN MESSAGE
4.7912		OPEN MESSAGE
4.7913		KEEPALIVE MESSAGE
4.8123		KEEPALIVE MESSAGE
4.8124		REQUEST MESSAGE
4.8344		REPLY MESSAGE
4.8345		CLOSE

Fig. 8. Wireshark output: PCEP captured packets

Fig. 8 shows the information of the captured packets at the PCEP session. The PCC is using IP address 130.206.162.XX, while PCE uses 150.244.56.XX. We can see that the PCE sessions are followed by our PCEP implementation:

- 1) **Session creation.** First there are two Open messages exchanged followed by Keepalive messages.

- 2) **Session Keepalive.** In this experiment, PCEP is working in intermittent mode. This is the reason why there are not Keepalive messages once the connection is established.
- 3) **Path Computation Process.** The Request message is sent from the PCC to the PCE, when a request arrives to the PCC. The PCE sends a Response message with the path computed based on the conditions included in the request message.
- 4) **Session close.** A Close message is sent to finish the PCEP connection.

Let us explain in detail the PCE response. Inside the Response message, there is a Explicit Route Object (ERO) with the route information. The route length in our scenario is four. These four hops are: the source IP address at UPNA (192.168.22.135), the incoming GRE IP address, the outgoing GRE IP address and the destination IP address at UAM (192.168.22.203). The PCE can select the GRE tunnel to FEDERICA or to PASITO. The first computed path is composed by 192.168.22.135, 192.168.22.131, 192.168.22.202 and 192.168.22.203. 192.168.22.131 and 192.168.22.202 are the FEDERICA GRE IP addresses at UPNA and UAM edge node respectively. The second path modifies the intermediates hops with 192.168.22.136 and 192.168.22.207, which are the the PASITO GRE IP addresses.

IV. CONCLUSIONS AND FUTURE WORK

MTE algorithms can help operators dealing with congestion in their IP layer networks. Moreover, these algorithms can increase network efficiency and reduce operational and ownership costs. This paper presents an implementation of a multi-layer technique in a realistic scenario involving three infrastructures: FEDERICA, PASITO and OneLab. The implementation allows transmitting the traffic over the most suitable path considering QoS restrictions. ARGOS cards are used to perform highly accurate network QoS measurements. Using high-accuracy equipment allows monitoring not only VoD services but more stringent services, such as real time applications. Furthermore, in our second experiment we have validated an implementation of the Path Computation Element Protocol in a network architecture lacking of a GMPLS control plane. This implementation allows the PCE to perform multi-layer engineering by selecting QoS constrained paths.

As future work, more complex MTE techniques will be analyzed with larger network topologies. We will analyze the scalability and the management of a high number of connections. The interconnection of FEDERICA with optical equipment would offer an IP over WDM scenario to test the effect of path establishment in the network.

ACKNOWLEDGMENT

This work has been partially funded by the Spanish Ministry of Education and Science under project ANFORA (TEC2009-

13385), by the Spanish Ministry of Industry, Tourism and Trade under PASITO project, and by the European Union under project OneLab2 (FP7-224263). Authors would like to thank Mauro Campanella (GARR, the project coordinator of FEDERICA) and Miguel Angel Sotos (RedIris) for their support to carry out this work.

REFERENCES

- [1] Cisco and Associates, "Hyperconnectivity and the Approaching Zettabyte Era." Cisco, Tech. Rep., 2010.
- [2] EU FP6-506760 IP IST NOBEL, "Deliverable D17: Preliminary Report on New Methods for Route Management (Intra and Inter-domain), and Accurate Statistical Models for Evaluating the Impact on QoS," Jan 2005.
- [3] J. Gabeiras, V. López, J. Aracil, J. Fernández Palacios, C. García Argos, O. González de Dios, F. Jiménez Chico, and J. Hernández, "Is Multi-layer Networking Feasible?" *Optical Switching and Networking*, vol. 6, no. 2, pp. 129 – 140, 2009.
- [4] B. Puype, Q. Yan, D. Colle, S. De Maesschalck, I. Lievens, M. Pickavet, and P. Demeester, "Multi-layer traffic engineering in data-centric optical networks," *Proceedings of Optical Networking Design and Modeling (ONDM)*, 2003.
- [5] H. Zhu, H. Zang, K. Zhu, and B. Mukherjee, "A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks," *Networking, IEEE/ACM Transactions on*, vol. 11, no. 2, pp. 285–299, 2003.
- [6] V. López, Ó. González de Dios, J. Hernández, R. Duque, C. García Argos, J. Jiménez, J. Fernández Palacios, and J. Aracil, "Performance evaluation of threshold-based multi-layer traffic engineering strategies," in *Proc. NOC/OC&I 2009*, April 2009.
- [7] S. Oliveira, O. González, E. Pérez, S. Andrés, and J. F. Palacios, "Techno-economic Approach for Multilayer Transport Networks," in *Networks and Optical Communications (NOC)*, June 2009.
- [8] S. Oliveira, E. Pérez, S. Andrés, L. Guijarro, and J. F. Palacios, "Economic Analysis for Transport Network Evolution," in *Conference of Telecommunication, Media and Internet Techno-Economics (CTTE)*, June 2009.
- [9] J.L. le Roux and J.P. Vasseur, "Path computation element (PCE) communication protocol," *IETF RFC 5440*, pp. 1–87, March 2009. <http://tools.ietf.org/html/rfc5440>.
- [10] M. N. Ellanti, S. S. Gorshe, L. G. Raman, and W. D. Grover, *Next Generation Transport Networks: Data, Management, and Control Planes*, 1st ed. Springer, April 2005.
- [11] "FEDERICA: Federated E-infrastructure Dedicated to European Researchers Innovating in Computing network Architectures," <http://www.fp7-federica.eu/>.
- [12] "PASITO: Plataforma de Análisis de Servicios de Telecomunicaciones," <http://www.rediris.es/proyectos/pasito/>.
- [13] "ONELAB FP7 Project," <http://www.onelab.eu/>.
- [14] V. Moreno, J. Garnica, F. Gomez-Arribas, S. Lopez-Buedo, I. Gonzalez, J. Aracil, M. Izal, E. M. na, and D. Morato, "High-accuracy network monitoring using ETOMIC testbed," in *7th EURO-NF Conference on Next Generation Internet*, June 2011.
- [15] K. S. Douglas Mauro, *Essential SNMP*, 2nd ed. O'Reilly Media, September 2005, vol. 1.
- [16] A. Farrel, J.P. Vasseur, and J. Ash, "A path computation element (PCE)-based architecture," *IETF RFC 4655*, pp. 1–40, August 2006. <http://tools.ietf.org/html/rfc4655>.
- [17] V. López, B. Huiszoon, Ó. González de Dios, J. Fernández Palacios, and J. Aracil, "Path Computation Element in Telecom Networks: Recent Developments and Standardization Activities." in *Optical Networking Design and Modeling (ONDM)*, February 2010.
- [18] E. Oki, T. Takeda, J.L. le Roux, and A. Farrel, "Framework for PCE-based inter-layer MPLS and GMPLS traffic engineering," *IETF RFC 5623*, pp. 1–34, September 2009. <http://tools.ietf.org/html/rfc5623>.