Dynamic Operation of an IP/MPLS-over-WDM Network Using an Open-Source Active Stateful BGP-LS-Enabled Multilayer PCE

Jose-Luis Izquierdo-Zaragoza[1], Jose-Juan Pedreno-Manresa[1], Pablo Pavon-Marino[1], Oscar Gonzalez de Dios[2], Victor Lopez[2]

[1] *Universidad Politécnica de Cartagena, Cuartel de Antiguones, Plaza del Hospital 1, 30202 Cartagena, Spain*

[2] *Telefónica I+D, C/ Don Ramón de la Cruz 82-84, 28006 Madrid, Spain*

e-mail: {josel.izquierdo, josej.pedreno, pablo.pavon}@upct.es, {ogondio, vlopez}@tid.es

ABSTRACT

Leveraging on reference implementations of PCEP and BGP-LS protocols, we have developed an active stateful multilayer PCE extension for the open-source Net2Plan planning tool to orchestrate multilayer service provisioning and restoration in IP/MPLS-over-WDM networks.

Keywords: PCE, BGP-LS, multilayer dynamic operation, IP-over-WDM, multilayer restoration, Net2Plan.

1. INTRODUCTION

Considering the operational complexity of multilayer networks, operators typically have two separated departments to manage the IP layer and the optical layer, with minimal interaction between them. Moreover, many processes require manual intervention, thus leading to a slow and very costly operation [1]. Fortunately, the increasing bandwidth demand along with dynamic (and flexible) service offerings are forcing carriers to explore solutions based on multilayer network orchestration [2], which are expected to optimize workflows for service provisioning, improve costs and to get rid from vendor lock-in.

In this work, we present an active stateful multilayer Path Computation Element (PCE) [3][4], where a centralized entity has the complete multilayer information to compute paths according to the network state (*stateful* feature), and it is able to issue network recommendations (i.e., lightpath rerouting upon failures) without any request (*active* feature) from the Path Computation Client (PCC). Exchange of requests and replies is based on IETF standards and drafts related to PCEP (PCE protocol) and BGP-LS [5], which are emerging as the de-facto solution in carrier-grade networks, using the `netphony-network-protocols` library [6]. Thanks to BGP-LS, we consider a real-world mechanism to automatically discover and update network state in the PCE.

The PCE is implemented as an external plugin of the open-source Net2Plan tool [7][8]. We test our implementation using an emulated multilayer network scenario (also developed as a plugin for Net2Plan). Our case study is based on typical multilayer workflows of operators, that is, service provisioning across several layers, and dynamic restoration under failures in the infrastructure. To the best of our knowledge, this is the first implementation of an active stateful PCE with multilayer capabilities.

The rest of the paper is organized as follows. Our reference testbed and multilayer scenario are described in Sections 2 and 3, respectively. Section 4 reports a case study on NSFNet network. Finally, we conclude in Section 5.

2. TESTBED SETUP

Fig. 1 depicts the reference architecture for the integrated multilayer PCE scenario. The physical network consists of IP/MPLS routers and optical equipment. As shown in Fig. 1, our testbed is based on the interplay of two Net2Plan (version 0.3.1) instances.

First, we implemented a basic multilayer network emulator able to maintain network state, as a tool inside the graphical user interface of Net2Plan (see Fig. 2a). This network is orchestrated by a controller managed from a Network Management System/Operations Support System (NMS/OSS), also emulated therein, from which the operator is able to monitor network state and establish new services, either (i) MPLS-TE paths (LSPs) or (ii) optical lightpaths. In this study, established MPLS services are assumed to be permanent. Also, we can create fiber failure/reparation events from the emulator.

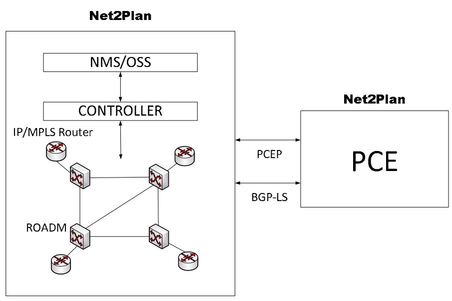
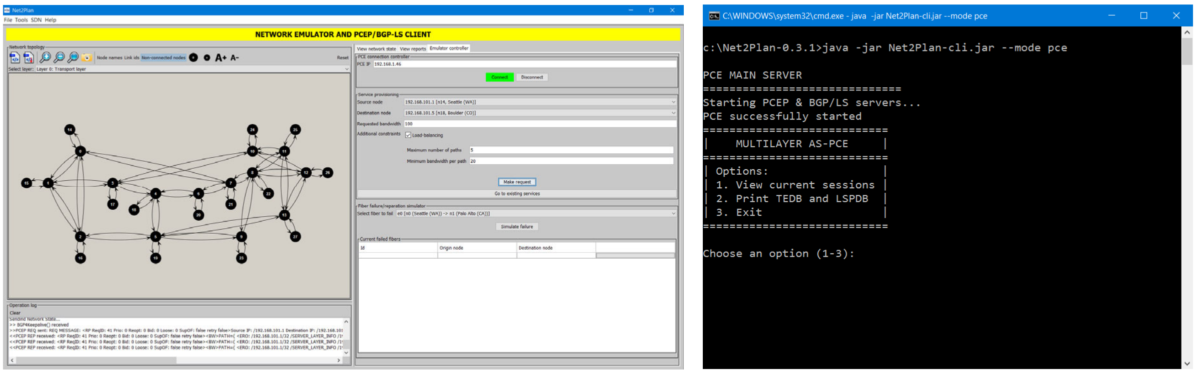


Figure 1. Reference architecture.

Second, the external PCE is in charge of receiving requests from the network emulator (acting as PCC), and giving the corresponding responses. In addition, it will be able to proactively react to failure/reparation notifications from the PCC, based on BGP-LS updates, suggesting changes (i.e., setup new lightpaths, rerouting MPLS services, and so on). This preliminary version runs as a daemon on top of Net2Plan command-line interface (see Fig. 2b).

Communication between PCC and PCE is uniquely based on PCEP and BGP/LS. To do this, the PCC establishes two sockets with the PCE. The first one is used for bidirectional PCEP communication [3]. In contrast, the second one is used for unidirectional BGP-LS interaction [5] (excluding Open and Keep-Alive messages) for BGP Updates, either for topology reporting or for failure/reparation notification, from PCC to PCE.



(a) Network emulator and PCEP/BGP-LS client (PCC) *(b) PCE daemon and hypervisor*

Figure 2. Snapshot of the implemented modules.

3. MULTILAYER DYNAMIC OPERATION

The primary challenge of active multilayer operation algorithms is to setup new LSPs in different layers while taking into account of different layer-specific constraints. These algorithms should be able to: (i) route lightpaths over the physical topology; and (ii) route connections in the lightpath-based client network. Last, but not least, they should be able to react to failure/reparation events, and to inform to the PCC of the corresponding actions to be performed. Thanks to the active and stateful nature of our multilayer PCE, and the "delegate" field [3] to support LSP delegation from the PCC to the PCE, the latter has full control over lightpaths and MPLS services. Namely, the PCE is able to suggest the establishment of new lightpaths and rerouting existing connections without an explicit request from the PCC. We consider a bottom-up strategy, that is, first we try to reroute affected lightpaths over the surviving physical topology, and then still-affected MPLS traffic is restored over the virtual topology.

4. CASE STUDY

In this section, we describe a reduced set of workflows, including message exchanges, from the implemented multilayer operation. We consider a multilayer version of the NSFNet reference topology. Optical fibers support up to 80 WDM channels working at 40 Gbps. We consider two instances of Net2Plan are running according to Section 2, whereas Wireshark packet-analyzer is used to monitor the PCC-PCE communication. This experiment can be replicated through a Net2Plan installation with the corresponding plugins in [9].

4.1 Workflow 1: Establishing PCC-PCE connection

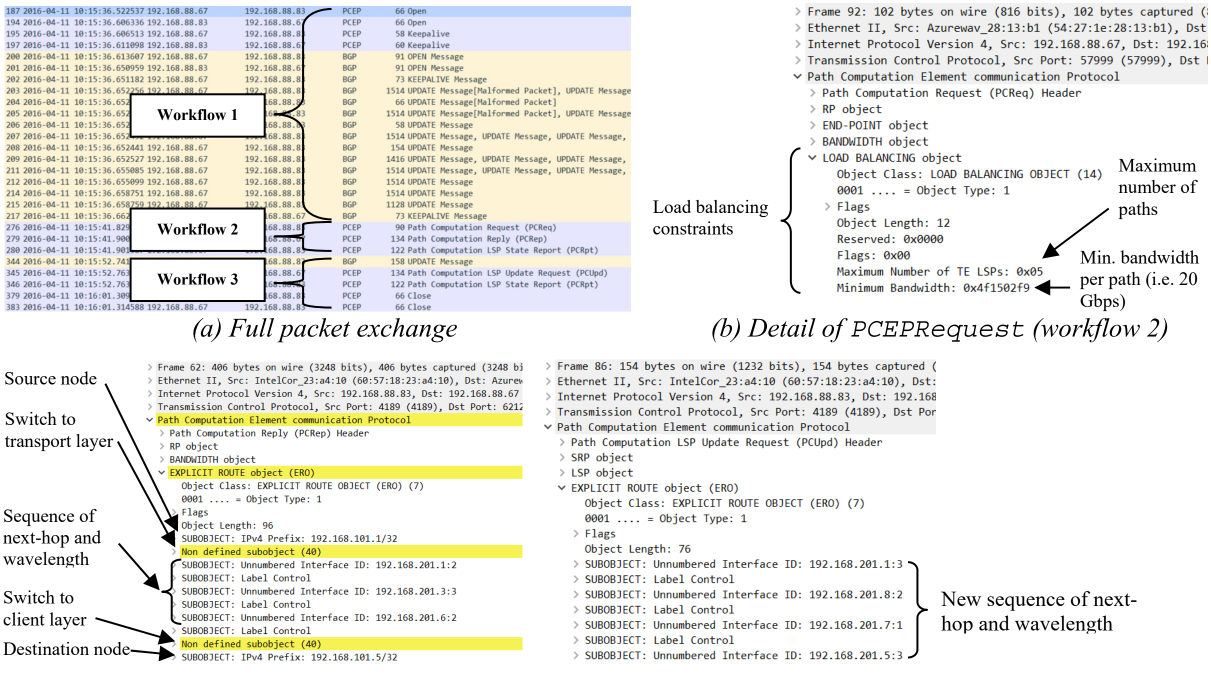
To establish the connection with the PCE, the PCC just needs to put its IP address in the tool, and click over "Connect" (see upper right side in Fig. 2a). Then, a sequence of PCEP/BGP Open and Keep-Alive messages is exchanged (frames 187-202, 217 in Fig. 3a), as defined by IETF documents. Upon connection, the PCC sends the whole topology information to the PCE via BGP-LS update messages using the Network Layer Reachability Information (NLRI) field [10] (frames 203-215 in Fig. 3a). Note that several `BGPUpdate` messages appear as "Malformed Packet". The reason is that Wireshark's BGP packet dissector does not reflect the latest BGP specification, at the time of writing, for topology information dissemination [5]. In any case, the used reference implementation of PCEP and BGP-LS protocols [6] has been reported to work properly [11], even interoperating with other implementations from other players, and validation processes using real-world equipment are being undergone.

4.2 Workflow 2: Service provisioning

Our NMS/OSS provides two different provisioning services to customers: (i) MPLS services (when "Client layer" is selected), where LSP provisioning is performed on top of the virtual topology, including automatic setup of new lightpaths to avoid blocking of LSP requests; and (ii) lightpath services (when "Transport layer" is selected),in order to allow setup of single lightpaths.

Focusing on the MPLS service provisioning, the connection might be established over multiple paths, if no one is able to carry the full connection. This is reflected by the `LOAD_BALANCING` (LB) object in the `PCEPRequest`. The workflow comprises the following steps (frames 276-280 in Fig. 3a):

1. The PCC issues a `PCEPRequest` (see Fig. 3b) for the MPLS service using the “Service provisioning” option (see middle section on the right side in Fig. 2a). Any requested bandwidth between 10 and 100 Gbps is valid (100 Gbps in the example). Optionally, it may include a LB object relaxing the routing policy to serve the LSP over several paths (see Fig. 3b).
2. The PCE looks for a sequence of lightpaths, or a set of them (if LB is allowed), with enough capacity to setup the LSP (using a capacitated shortest path algorithm). If fails, try to establish an end-to-end, or several (if LB is allowed) lightpath to carry the whole connection. The algorithm uses a shortest path routing with first-fit for the wavelength assignment.
3. If a path (or more than one, if LB is allowed) could not be found, the PCE issues a `NO_PATH` reply and the request is withdrawn. Otherwise, go to 4.
4. The PCE issues a `PCEPReply` (see Fig. 3c) with the allocation found. If multiple paths exist (and LB is allowed), the `PCEPReply` will include multiple Explicit Route Objects (EROs). In case a path requires a new lightpath, this information is handled between a pair of `SERVER_LAYER` objects in the ERO.
5. The PCC then updates the network state, and issues a `PCEPReport` message with “delegate” option. Hence, from now on the PCE has control over lightpaths and MPLS services (i.e. failures will be proactively handled by the PCE).



(a) Full packet exchange *(b) Detail of `PCEPRequest` (workflow 2)*

(c) Detail of `PCEPReply` (workflow 2) *(d) Detail of `PCEPUpdate` (workflow 3)*

Figure 3. Traces from Wireshark.

4.3 Workflow 3: Reaction to fiber link failure

From the emulated NMS/OSS, the PCC can emulate a fiber link failure to evaluate the reaction of our multilayer PCE algorithm using the “Simulate failure” option (see lower right side in Fig. 2a). We would like to remark that the PCC sends topology updates via BGP-LS and the PCE automatically reacts in case of failure, without the need for an explicit `PCEPRequest` message.

The workflow is as follows (frames 344-346 in Fig. 3a):

1. The PCC issues the failure via a `BGPUpdate` message with the `MP_UNREACH` path attribute corresponding to the failed fiber indicated by the NLRI.
2. The PCE tries to reroute affected lightpaths over the surviving topology. In case some lightpath could not be rerouted, then we try to restore MPLS services looking for a sequence of lightpaths (or set of them) with enough capacity to carry the affected routes.
3. The PCE sends back to the PCC a list of ordered actions in a `PCEPUpdate` (see Fig. 3d) message to be performed, with “delegate” option.
4. The PCC acknowledges with `PCEPReport` messages, again with “delegate” option.

In Fig. 4 we represent the effect of lightpath rerouting upon fiber failure. In our example, a failure in fiber 2-5 (red dashed line) triggers rerouting of the selected lightpath from 0 to 5 (original and rerouted fiber paths are highlighted with thick solid and dashed lines, respectively).

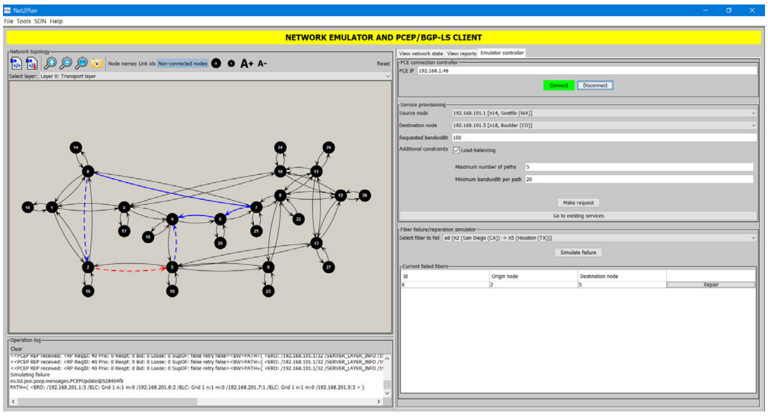


Figure 4. Lightpath rerouting upon fiber failure.

Finally, in our case study we do not consider any "reverting back to normal" mechanism, where lightpaths rerouted during failure states are reverted to their original state upon reparation, in order to avoid traffic outage during lightpath migration [12]. We can actually simulate the reparation of the failed fiber link using the "Simulate reparation" option, which sends a BGP-LS reparation event (`BGPUpdate` with `MP_REACH` attribute), but no further action will be performed by the PCE.

5. CONCLUSIONS AND FURTHER WORK

This work demonstrates an active stateful multilayer PCE with full topological information, based on the open-source Net2Plan tool, able to intercommunicate with the network using PCEP and BGP-LS protocols. Thanks to the usage of standard communication protocols, we should be able to substitute our emulator with real equipment. We also plan to extend our work following the concept of "in-operation" network planning [13]. In other words, we will employ capabilities of our active stateful multilayer PCE to explore i.e. lightpath reoptimization mechanisms.

ACKNOWLEDGEMENTS

This work was partially supported by the FPU fellowship program of the Spanish Ministry of Education, Culture and Sports (ref. no FPU14/04227) and by the Spanish project grants TEC2014-53071-C3-1-P (ONOFRE) and TEC2015-71932-REDT (ELASTIC).

REFERENCES

[1] J.-P. Fernandez-Palacios, *et al.*: IP and Optical Convergence: Use Cases and Technical Requirements, White Paper, Jan. 2014.
[2] O. Gerstel, V. Lopez: The Need for SDN in Orchestration of IP over Optical Multi-Vendor Networks, in *Proc. ECOC 2015*, Valencia, Spain, Sep.-Oct. 2015.
[3] J.-P. Vasseur, J.-L. Le Roux: Path Computation Element (PCE) Communication Protocol (PCEP), RFC 5440, Mar. 2009.
[4] E. Crabbe, *et al.*: PCEP Extensions for Stateful PCE, draft-ietf-pce-stateful-pce-14, Mar. 2016
[5] H. Gredler, *et al.*: North-Bound Distribution of Link-State and Traffic Engineering (TE) Information using BGP, RFC 7752, Mar. 2016.
[6] Java Library of Networking Protocols. [Online] Available: https://github.com/telefonicaid/netphony-network-protocols [Last accessed: Apr. 2016]
[7] Net2Plan – The Open-Source Network Planner. [Online] Available: http://www.net2plan.com [Last accessed: Apr. 2016]
[8] P. Pavon-Marino, J.L. Izquierdo-Zaragoza: Net2Plan: An open-source network planning tool for bridging the gap between academia and industry, *IEEE Network*, vol. 29, no. 5, pp. 90-96, Sep.-Oct. 2015.
[9] Repository of Net2Plan Plugins for Active Stateful Multilayer PCE, Multilayer Network Emulator and PCEP/BGP-LS Client. [Online] Available: https://github.com/girtel/n2p-plugins-activePCE [Last accessed: Apr. 2016]
[10] T. Bates, *et al.*: Multiprotocol Extensions for BGP-4, RFC 4760, Jan. 2007.
[11] H. Gredler, *et al.*: BGP Link-State Information Distribution Implementation Report, draft-ietf-idr-ls-distribution-impl-04, May 2015.
[12] A. Mayoral, *et al.*: Minimizing Resource Protection in IP Over WDM Networks: Multi-layer Shared Backup Router [Invited], *IEEE/OSA JOCN*, vol. 7, no. 3, pp. A440-A446, Mar. 2015.
[13] L. Velasco, *et al.*: In-Operation Network Planning, *IEEE Com. Mag.*, vol. 52, no. 1, pp. 52-60, Jan. 2014.