# First Multi-partner Demonstration of BGP-LS enabled Inter-domain EON control with H-PCE

**O. González de Dios[1], R. Casellas[2], R.Morro[3], F. Paolucci[4],**
**V. López[1], R. Martínez[2], R. Muñoz[2], R. Vilalta[2], P. Castoldi[4]**
[1] *Telefónica I+D/GCTO, Distrito Telefonica, 28050 Madrid (Spain). Contact email: oscar.gonzalezdedios@telefonica.com*
[2] *CTTC. Av. Carl Friedrich Gauss n7, 08860 Castelldefels, Barcelona*
[3] *Telecom Italia, Via G. Reiss Romoli 274, 10148 Torino, Itay*
[4] *CNIT-Scuola Superiore Sant'Anna, Pisa, Italy*

**Abstract:** The control of Multi-domain Elastic Optical Networks (EON) is possible by combining H-PCE based computation, BGP-LS topology discovery, remote Instantiation via PCEP, and signaling via RSVP-TE. This paper presents the first multi-platform demonstration that fully validates such control architecture achieving multi-protocol interoperability.

**OCIS codes:** (060.0060) Fiber optics and optical communications; (060.4250) Networks.

## 1. Introduction

Elastic Optical Networks (EON) represent the state-of-the-art of connection-oriented optical networks. Thanks to the recent advances in the design of flexible bandwidth-variable transponders (BVT), capable of transmit/receive signals with configurable physical parameters (i.e., bitrate, modulation format), and to the availability of Spectrum Selective Switches (SSS), capable of switching frequency slices multiple of 12.5GHz, EON have the potential to enable fully configurable multi-bitrate lightpaths thus increasing service-oriented flexibility and overall network capacity [1].

Adequate control plane operation is required to achieve lightpath provisioning, along with additional EON-specific procedures, such as elastic operation and re-optimization (e.g., defragmentation). In the context of the ABNO architecture [2], a centralized multi-component controller is envisioned and the Path Computation Element (PCE) may be used, besides path computation, as functional component for direct lightpath instantiation and release [3]. In the case of multi-domain networks, hierarchical approach is considered for path computation. Hierarchical PCE (HPCE), comprising a parent PCE (pPCE) and several per-domain Child PCEs (cPCE), has been successfully demonstrated for WSON [4]. However, both optimal domain sequence and intra-domain segment selection could be achieved only if pPCE is aware of detailed Traffic Engineering (TE) topology. Such information may be available at pPCE without scalability issues by resorting to the recently proposed TE Link State Information extensions to BGP (BGP-LS) [5], as experimentally demonstrated in [6].

In this paper, an extended HPCE architecture is considered to control multi-domain EON based on GMPLS. Routing (OSPF-TE and BGP-LS), path computation/instantiation (PCEP) and signaling (RSVP-TE) protocol extensions necessary to enable inter-domain EON lightpath provisioning are detailed. For the first time, such extensions are evaluated in a complete distributed multi-partner control plane testbed in order to fully validate inter-operability among multi-platform and/or multi-vendor network/node controllers. Complete provisioning performances are detailed, including BGP-LS topology update, path computation (i.e., segment computation, selection and path concatenation, including BVT end-point indication and configuration), instantiation and end-to-end RSVP-TE signaling.

## 2. Architecture

The propose architecture is built around the hierarchical PCE (HPCE) framework, in which a parent PCE (pPCE)

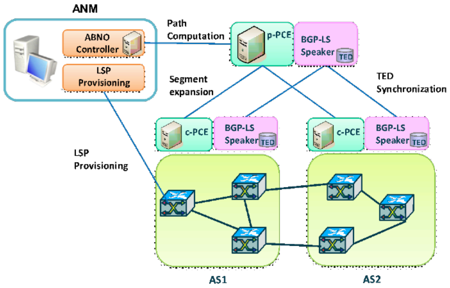


Fig. 1. Control architecture

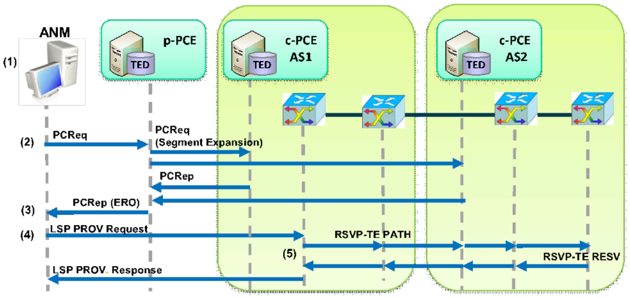


Fig. 2. Message Flow.

coordinates several children PCEs (cPCE), one per network domain (Figure 1). The pPCE is in charge of domain selection and inter-domain path computation. Children PCEs are responsible for segment expansion, i.e. for path computation in their respective domains. BGP-LS is deployed for topology abstraction and to export inter-domain TE information to the pPCE. LSP provisioning is triggered by an SDN controller that monitors network resources utilizations and is able to decide the optimal network configuration based on the status, bandwidth availability and user service. It leverages recently proposed extensions to the PCEP protocol for the so-called stateful and active PCE and includes instantiation capabilities. A single end-to-end RSVP-TE signaling session is used for setting up the connection between the domains; this allows simplified setup and teardown procedures, especially in case of exceptions handling, w.r.t the case of separated signaling sessions (one per domain).

## 3. Control Plane procedures and protocol extensions

The designed and implemented control plane is based on the GMPLS/PCE architecture and protocols, augmented with the use of BGP-LS as a topology dissemination protocol and PCEP with stateful and instantiation extensions as the selected instantiation interface (although not necessarily used by a PCE, but by a provisioning manager).

Control Plane Procedures: The control plane covers several functions such as automatic topology discovery and management; path computation, and dynamic Label Switched Path (LSP) provisioning. The path computation function is based on the HPCE [7] architecture and PCEP [3] is used as the provisioning protocol. For scalability and confidentiality, the pPCE operates on an discovered abstract topology, where BGP-LS [5] is used by dedicated entities at each domain (e.g. collocated with the child PCE) as the protocol for topology and reachability advertisement, allowing flexibility in what is announced, and under operator policy. The pPCE can thus construct a multi-domain topology including, notably inter-domain connectivity. Let us detail the process with the help of Figure 2. Upon request (1), the SDN controller triggers the provisioning. First, it requests a multi-domain path computation (2), which is a two-step process in which the pPCE obtains the domain sequence and then requests the children to expand the domain path within their respective domains. The pPCE composes and end-to-end ERO, which, by default, uses unnumbered interfaces represent outgoing TE links but, to convey information about the ingress port, it can be prepended with an additional ingress interface (facing the client) and may either end with a IPv4 prefix address or an unnumbered interface meaning that the LSP ends at the output interface with an additional cross-connect. Explicit label control (ELC) conveys information about the outgoing label (frequency slot) that will be used by the downstream node in switching. The actual LSP provisioning takes place after the end to end path has been computed (3), and the provisioning manager uses the PCEP interface with the ingress node to request a Path establishment (4). It is based on the use of PCInitiate and the PCRpt messages: the PCInitiate includes the SRP, LSP, ENDPOINTS, ERO objects, and instructs the ingress node to initiate the signaling procedure, based on the Path/Resv RSVP-TE message exchange with an end-to-end session (5). Upon completion of the signaling process, the PCRpt message is sent back to the provisioning manager, additionally including the route object RRO and the allocated frequency slot.

Protocol Extensions: the inherent aspects of flexi-grid networks require extending all involved protocols. Common to the extensions are, notably, a new label format based on 64 bit encoding of the central frequency and slot width and the ability to disseminate the status of the nominal central frequencies on a per link basis, using a bitmap format encoding, and used uniformly in the PCEP, BGPLS, OSPF-TE and RSVP-TE protocols. Within a domain, OSPF-TE is extended to disseminate, as a minimum, the status of each link in terms of optical spectrum [8], detailed as a TLV within the Interface Switching Capability Descriptor (ISCD) of the Link TLV in a TE Link State Advertisement (LSA). In RSVP-TE, the sender traffic parameters or SENDER_TSPEC (included in the sender descriptor of the Path message) as well as the FLOWSPEC object in the flow descriptor (Resv message) have new types to convey the desired and assigned frequency slot width, respectively [9]. Likewise, PCEP requires GMPLS extensions [10], notably for the support of generalized bandwidth and endpoints and new objects including suggested and assigned labels and bitmap sets to enable Routing and Spectrum Assignment (RSA).

## 4. Experimental Demonstration in the multi-partner testbed

Idealist Multi-partner Test-bed: the Idealist Multi-partner Control Plane Test-bed interconnects four European research institutions, located in Madrid (Telefónica I+D), Barcelona (CTTC), Torino (TI) and Pisa (CNIT). The testbed physical topology is depicted in Figure 3. Partners' premises are connected (at the control plane level) by means of dedicated IPsec tunnels. The resulting low level connectivity layout is a hub, centered at CTTC. Static routing entries provide full connectivity between partners' private addresses, secured and isolated from the rest of Internet traffic. On top of this distributed control plane connectivity network, logical relationships between PCEs are established, in particular between Telefónica I+D PCE, acting as pPCE, and the other PCEs, acting as cPCE, as

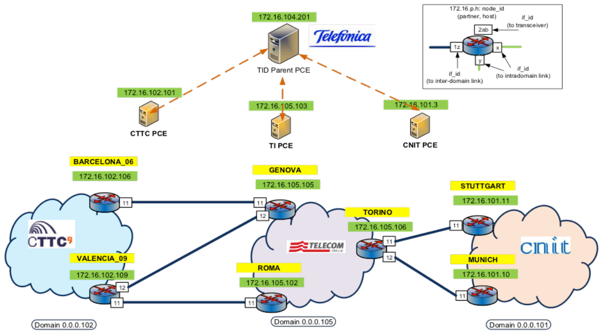


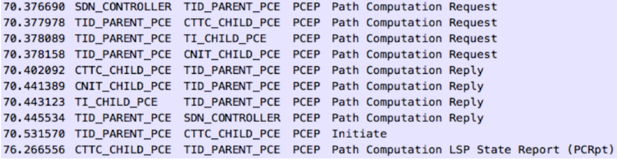
Fig. 3. Multi-partner European testbed topology

| Ping times | |
|---|---|
| | avg/mdev |
| TID-CTTC | 14.3/1.454 ms |
| TID-TI | 65.228/3.600 ms |
| TID-CNIT | 63.768/8.397 ms |
| CTTC-TI | 51.31/1.997 ms |
| TI-CNIT | 100.11/3.502 ms |

| Computing times | |
|---|---|
| | avg |
| TOTAL Comp | 65,5 ms |
| TID-CTTC | 24,1 ms |
| TID-TI | 65 ms |
| TID-CNIT | 63,2 ms |

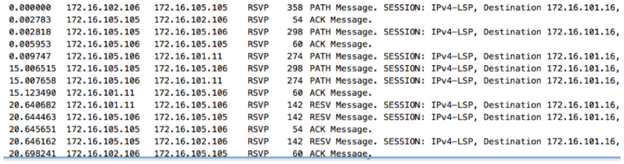
| setup time | |
|---|---|
| | avg/mdev |
| CTTC-TI-CNIT | 5,6/0,32 sec |

Fig. 4. Performance Results



```
70.376690  SDN_CONTROLLER  TID_PARENT_PCE  PCEP  Path Computation Request
70.377978  TID_PARENT_PCE  CTTC_CHILD_PCE  PCEP  Path Computation Request
70.378089  TID_PARENT_PCE  TI_CHILD_PCE    PCEP  Path Computation Request
70.378158  TID_PARENT_PCE  CNIT_CHILD_PCE  PCEP  Path Computation Request
70.402092  CTTC_CHILD_PCE  TID_PARENT_PCE  PCEP  Path Computation Reply
70.441389  CNIT_CHILD_PCE  TID_PARENT_PCE  PCEP  Path Computation Reply
70.443123  TI_CHILD_PCE    TID_PARENT_PCE  PCEP  Path Computation Reply
70.445534  TID_PARENT_PCE  SDN_CONTROLLER  PCEP  Path Computation Reply
70.531570  TID_PARENT_PCE  CTTC_CHILD_PCE  PCEP  Initiate
76.266556  CTTC_CHILD_PCE  TID_PARENT_PCE  PCEP  Path Computation LSP State Report (PCRpt)
```

Fig. 5 PCEP capture of lightpath provisioning at TID pPCE



```
0.000000   172.16.102.106  172.16.105.105   RSVP   358 PATH Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
0.002783   172.16.105.105  172.16.102.106   RSVP    54 ACK Message.
0.002818   172.16.105.105  172.16.105.106   RSVP   298 PATH Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
0.005953   172.16.105.106  172.16.105.105   RSVP    60 ACK Message.
0.009747   172.16.105.106  172.16.101.11    RSVP   274 PATH Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
15.006515  172.16.105.105  172.16.105.106   RSVP   298 PATH Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
15.007658  172.16.105.106  172.16.101.11    RSVP   274 PATH Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
15.123490  172.16.101.11   172.16.105.106   RSVP    60 ACK Message.
20.640682  172.16.101.11   172.16.105.106   RSVP   142 RESV Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
20.644463  172.16.105.106  172.16.105.105   RSVP   142 RESV Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
20.645651  172.16.105.105  172.16.105.106   RSVP    54 ACK Message.
20.646162  172.16.105.105  172.16.102.106   RSVP   142 RESV Message. SESSION: IPv4-LSP, Destination 172.16.101.16,
20.698241  172.16.102.106  172.16.105.105   RSVP    60 ACK Message.
```

Fig. 6 RSVP-TE capture at Telecom Italia border node "Torino".

shown in Figure 3. The PCEs of the testbed have been independently developed by each partner. BGP-LS speakers are implanted by each partner.

Telefónica I+D pPCE is a multi-threaded application developed in Java 1.6. It accepts sessions from cPCEs, maintaining each session with a specific thread which handles all the messages exchange. Also, a dedicated thread is used for each BGP-LS session, building the multi-domain TE Database (TED), in which the nodes are domains, and the edges are the inter-domain links, and the reachability information is obtained by node advertisements. CNIT testbed comprises 7 C++-based EON controllers capable of dynamically configuring co-located SSS by means of USB interface and a C++ based cPCE performing advanced impairment-aware computation. CTTC domain emulates a 14 node mesh network that represents a national (spanish) photonic transport segment and a stateful cPCE. TI test-bed is composed of 6 Linux boxes running GMPLS-based control plane processes, whose architecture is in line with Rec. ITU-T G.8080 in terms of architectural components, emulating a single network node each.

*Performance Evaluation*: first of all, the architecture has been validated functionally, and the interoperability of four PCEP implementations has been achieved. Figure 5 shows a Wireshark capture in the parent PCE machine, showing the path computation interactions and the later initiation procedure. Figure 6 shows the RSVP-TE interactions that are triggered by the PCEP instantiation message. The experiment shows that the total computation time, including all the interactions between parent and child PCEs is 68 ms. The initiation time is 5.7 seconds. Most of the time of the initiation is due to the configuration of the real OXC WSS based flexi-grid nodes in CNIT testbed. It has to be taken into account that both CTTC and TI test-beds run on emulated nodes, and thus set-up time is faster. The latencies between the different components is shown in figure 4, that shows the mean times of the important steps as well.

## 5. Conclusions

Interoperability between four implementations of BGP-LS and PCEP protocols, extended to support GMPLS and EON for path computation and instantiation, and between three of extended RSVP-TE, has been successfully demonstrated. Results show that configuring the nodes accounts for most of the time in the provisioning process.

**Acknowledgements**
The research is partially funded by the European Community's in the IDEALIST project (317999).

## 6. References

[1] M. Bohn et al.," Elastic Optical Networks: the Vision of the ICT Project IDEALIST", FuNeMS 2013.
[2] D. King and A. Farrel. A PCE-based Architecture for Application-based Network Operations. draft-farrkingel-pce-abno-architecture-13
[3] E. Crabbe et al., "PCEP extensions for PCE-initiated LSP setup in a stateful PCE model," IETF draft-ietf-pce-pce-initiated-lsp-01
[4] F. Paolucci et al., "Experimenting Hierarchical PCE architecture in a distributed multi-platform control plane testbed", OM3G.3, OFC 2012.
[5] H.Gredler et al., "North-Bound Distribution of Link-State and TE Information using BGP", IETF draft-ietf-idr-ls-distribution-06 ..
[6] M. Cuaresma et al. "Experimental Demonstration of H-PCE with BPG-LS in elastic optical networks", ECOC 2013
[7] D. King and A. Farrel, RFC 6805
[8] F. Zhang et al. GMPLS OSPF-TE Extensions in support of Flexible Grid in DWDM Networks
[9] F. Zhang et al. RSVP-TE Signaling Extensions in support of Flexible Grid draft-ietf-ccamp-flexible-grid-rsvp-te-ext-00
[10] C. Margaria et al. PCEP extensions for GMPLS draft-ietf-pce-gmpls-pcep-extensions-10