IDEALIST Control Plane Architecture for Multi-domain Flexi-Grid Optical Networks

Ramon Casellas*, Raul Muñoz*, Ricardo Martínez*, Ricard Vilalta*, Filippo Cugini†, Francesco Paolucci†,
Oscar González‡, Victor López‡, Juan Pedro Fernández-Palacios‡,
Roberto Morro§, Andrea Di Giglio§, Daniel King¶, and Adrian Farrel¶

*CTTC, Av. Carl Friedrich Gauss 7, Castelldefels, 08860, Spain
†CNIT, Pisa, Italy
‡Telefónica I+D, c/ Don Ramón de la Cruz 84, Madrid, 28006, Spain
§Telecom Italia, Via Reiss Romoli 274, Torino, 10148, Italy
¶Old Dog consulting, Llangollen, LL20 8EB, UK

***Abstract*—The ICT IDEALIST project has designed and implemented a GMPLS/PCE control plane and provided significant flexi-grid DWDM network input for IETF standardization. This paper highlights the multi-domain considerations, challenges, and how these were solved. The paper objectives are two-fold: to provide an overview of the major issues and challenges when designing the aforementioned control plane and, second, to summarize the main architectural, functional and protocol choices. The control plane architecture is based on the concept of abstracted network layer and using hierarchical, stateful PCE, in which the parent PCE coordinates the domain selection and multi-domain provisioning, while delegating segment expansion and intra-domain provisioning to the corresponding child PCE: it guarantees an end-to-end network service by concatenating Label Switched Paths at each domain, which are effectively set up independently by the underlying GMPLS control plane. The BGP-LS protocol is used for northbound link state distribution, allowing the parent to construct an abstract topology of the underlying network.**

I. Introduction

Optical Transport Networks (OTN), as defined in ITU-T Recommendation G.872 [1], are composed of optical network elements connected by optical fibers, able to provide the functionality of transport, multiplexing, routing, management, supervision and survivability of optical channels (OCh) carrying client signals, constrained by a fixed ITU-T DWDM grid [2]. Within the IETF CCAMP working group, such networks are referred to as Wavelength Switched Optical Networks (WSON) and a set of normative documents are being published to use GMPLS for the automatic establishment of the so-called *lightpaths*.

However, such fixed grid is not adapted to high data rates, and is inefficient when a wavelength is assigned to a low rate optical signal (e.g., 10 Gb/s) that does not fill the entire wavelength capacity. To overcome this major limitation, WSONs are evolving towards Spectrum Switched Optical Networks (SSON) using a flexible grid, in which the optical spectrum is characterized by a frequency grid having Nominal Central Frequencies (NCF) with a lower spacing granularity (6.25 GHz) and the required amount of optical bandwidth for an optical channel can be dynamically and adaptively allocated, in multiples of a given slot width granularity (12.5 GHz), determined by the signal modulation format and its data rate. In the next section, we briefly mention the main components of control plane within a single Traffic Engineering (TE) domain. The applicability in multi-domain networks constitutes the main contribution of this paper.

II. Overview of the Control Plane architecture and protocols for a Single domain

The control plane being designed within IDEALIST is based on the GMPLS/PCE framework and protocols. IDEALIST current scope is the control of the media layer [1]: a network media channel transports a single Optical Tributary Signal, or OTS (a particular example of OTS is the Optical Channel Payload, or OCh-P). The main requirement of the control plane is the dynamic establishment and release of flexi-grid Label Switched Paths (LSPs), representing a media channel, which are switched in media channel matrixes (cfr. Figure 1). GMPLS labels locally represent the media channel and its associated *frequency slot*. Network media channels are considered a particular case of media channels when the end points are transceivers.

A. Resource Discovery and Topology Dissemination

The Open Shortest Path First with Traffic Engineering Extensions (OSPF-TE) routing protocol is being extended to support the dissemination, via Link and Node Link State Advertisements (LSAs), of TE attributes that enable the building of a network topology view, commonly referred to as the Traffic Engineering Database (TED). Thus, the control plane needs to have a model of all the switching elements and their restrictions (e.g., devices may have a different minimum slot size or cannot support all sizes). The TED is used as an input in the path computation function. Although such computation can be deployed directly in the GMPLS controllers that constitute the control plane, different considerations such as the specifics of the optical layer technology or the benefits of a stateful Path Computation Elements (PCE) justify the choice for their deployment. This does not preclude the use of source based path computation or hybrid approaches combining PCE-based provisioning and source-based recovery.

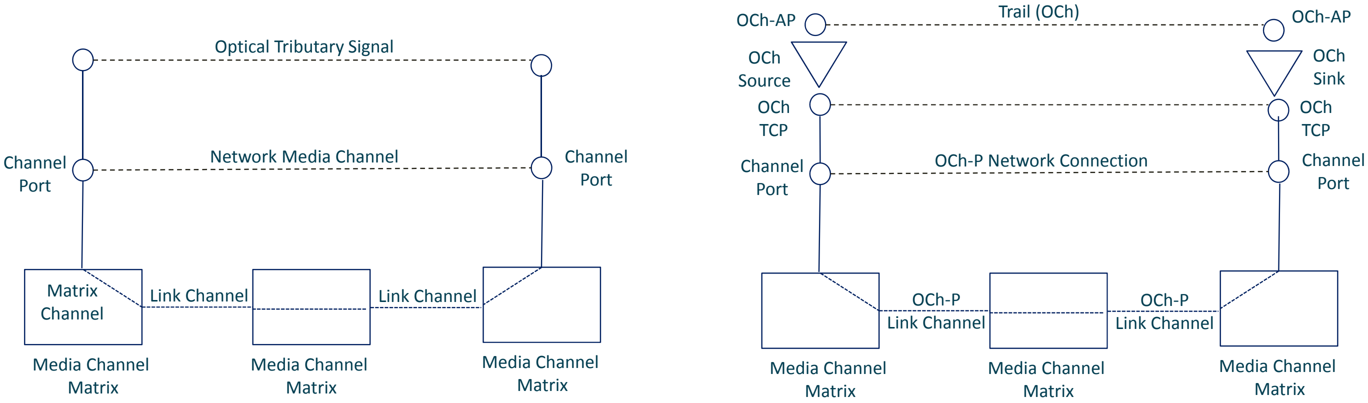

Fig. 1. IDEALIST current scope deals with the establishment of LSPs that represent (network) media channels. network media channel transports an Optical Tributary Signal, OTS (left). A particular case corresponds to the OCh-P (right)

The TED needs to capture not only the current status of the TE links (in terms of available frequency slots, and the overall status of the optical spectrum on a per fiber basis), but also the fact that optical transmitters/receivers may have different tunability constraints, and that media channel matrixes may have switching restrictions and/or highly asymmetric switching capabilities [3]. Each of these constraints is being addressed by extending the Interface Switching Capability Descriptor (ISCD) associated to each link port. The specific encoding of such information is still an open issue: possible encodings are based on either describing the status of the nominal central frequencies, thus favoring the choice of a bitmap encoding, or describing available ranges slots with an inclusive or exclusive frequency slot list. Tunability granularity, both in terms of selectable nominal frequencies and slot widths, need also to be disseminated on an interface-basis. By design, we aim at reusing, where appropriate, existing normative documents that were or are being published in the scope of WSON.

B. *Stateful Path Computation*

The PCE architecture was proposed to provide effective constraint-based path computations. So far, the PCE has been mainly deployed with a stateless architecture, i.e. the PCE only relies on the TED which includes information on resource utilization. More recently, the PCE architecture has been extended with stateful capabilities, enabling the attributes of the established LSPs (e.g., the route) to be stored and maintained at the LSP State Database (LSPDB) [4]. Furthermore, a stateful PCE may also include the active functionality which enables the PCE to issue recommendations to the network, e.g. to dynamically update LSP parameters through the PCE Communication Protocol (PCEP). In IDEALIST, the (active) stateful architecture has been adopted to enable a number of advanced traffic engineering functionalities, including elastic LSP operations and global defragmentation in flexi-grid networks. For example, the PCE is able to account for the actual network conditions, run complex re-optimization algorithms, and operate on existing LSPs to reduce the overall network fragmentation. The implementation of the stateful functionality has also to account for some deployment considerations, mainly related to reliability, synchronization (e.g., after restart) and scalability issues. In terms of scalability, the stateful PCE is not designed to be operated over the entire Internet. On the contrary, its domain of visibility has to be adequately dimensioned, considering a sufficiently over-provisioned system.

C. *Signaling aspects and protocol extensions*

The signaling mechanism within IDEALIST control plane aims at allowing the dynamic provisioning, re-routing and recovery of network media channels. A basic building block is the newly proposed 64-bit label encoding, which extends the 32-bit label that identifies a wavelength in WSON with the information regarding the slot width. The proposed approach extends RSVP-TE with a new switching type for SSON and adds new types for both the sender descriptor traffic specification object, conveying the requested slot width, and for the flow descriptor that conveys the allocated slot width (cfr. [5]). The use of existing procedures for the dynamic rerouting of connections is being addressed. Initial considerations involve the establishment of LSPs with a Shared Explicit (SE) reservation style, which allows modification of connection attributes such as its central frequency or its slot width.

Finally, one key point is that a set of media channels can be used to transport signals that have a logical association between them and are not necessarily adjacent in frequency. Thus, the control plane architecture allows multiple media channels to be logically associated and allows the co-routing of a set of media channels logically associated.

III. Control Plane architecture for a Multi-Domain Network

Even when under the control of a single administrative entity, transport networks may be segmented for technical or scalability reasons (e.g., in the form of vendor islands). Such multi-domain networks are characterized by the fact that no single entity has full topology (TE) visibility, affecting optimality and efficient resource usage. In IDEALIST, we rely on a hierarchical PCE (H-PCE) approach, scoped in the framework of what we refer to as *interconnected traffic engineered networks*, as detailed in Section III-A

The macroscopic functional and protocol architecture is shown in Figure 2. One basic assumption that we make here is that the domains interconnection is done by means of ”border links” rather than ”border nodes”. The former is the case when two devices, one per each domain, are interconnected by a shared link, while the latter refers to the case where a single network element belongs to both domains. We assume that an Adaptive Network Manager (ANM) triggers, by means of a provisioning interface towards the parent PCE (pPCE), the activation of network connectivity services, which maps to the actual establishment and release, via de control plane, of elastic connections. The actual provisioning of the connection is coordinated by the pPCE and ultimately delegated to the underlying GMPLS control plane at each domain.

A. Interconnected Traffic Engineered Networks

As highlighted earlier, TE networks are typically segmented into Interior Gateway Protocol (IGP) domains, with TE information being contained within each domain to ensure network scaling and confidentiality. However, some client services would benefit from an end-to-end TE path across a number of connected domains, therefore it would be beneficial to expose a limited amount of TE information about each domain, to help with the modeling, computation and establishment of end-to-end TE services across multi domain networks. The concept of TE reachability has been defined in [6] and may be categorized by TE attributes such as: TE metrics, hop count, available bandwidth, delay, shared risk, etc. A summary, or subset, of TE reachability information should be provided from each domain so that a client node, PCE or ANM can determine whether they can establish a TE path from, across and to another domain, with the required TE metrics.

In order to compute a path across the transport (server layer) using the TE reachability attributes of the source domain, candidate transit domains, and the destination domain, the TE reachability knowledge of each domain must be instantiated. This is achieved via TE network abstraction, which is the synthesizing of reported TE attribute information for each domain and inter-domain link. This provides the aggregated TE reachability information and subsequent abstracted topology representation, known as virtual links and nodes (virtual topology). This transport network abstraction and creation of a virtual topology does not represent all possible connectivity options, but instead provides a connectivity matrix based on current TE attributes that are being reported from within each domain. While abstraction uses available TE information, it may also be subjected to network policy and management choices. Thus, not all potential connectivity would be advertised.

If the current transport connectivity does not meet required or expected client demands new peer connections can be established. These TE LSP tunnels will span the transport domains used to achieve the required or expected connectivity. These transit LSPs are the key building blocks of the end-to-end connectivity for the client. It is expected that planning will be required to ensure the required connectivity is available, but dynamic or on-demand requests could be supported, but should be subject to policy considerations. Once a suitable topology exists it can be abstracted into a virtual topology provided as a TED to the client node, ANM or PCE for computing end-to-end TE-based services across the multi domain network without exposing the internal domain topologies and exhaustive TE information of the transport network.

B. Multi-domain Topology Management and Inter-domain routing

In IDEALIST proposed hierarchical architecture, it is required to maintain a domain topology map at the pPCE, representing a view of the child domains and their interconnectivity, e.g. its *abstract representation*. The procedure and protocol mechanisms for disseminating and constructing of the pPCE TED may be provided using a number of mechanisms, currently being evaluated within the project. It is important to note that all mechanisms are subject to policy within their originating domains.

- The pPCE could joining the IGP instance of each child PCE domain. The attributes of the interdomain links may be distributed within a domain by TE extensions to the IGP, as in [7]. However, it would break the domain confidentiality principles and it is subject to scalability issues. Alternatively, [8] points out that in ASON models it is possible to consider a separate instance of an IGP running within the parent domain with the participation of the child PCEs.
- Authors in [9] proposed the embedding in PCEP Notifications both intra domain and inter-domain LSAs. This approach was experimentally demonstrated in a multi-partner testbed [10]. However, it is argued that the utilization of PCEP is beyond the scope of such protocol.
- Use north-bound distribution of TE information, by means of the BGP-LS protocol [11]. With this approach, there is a BGP speaker in each domains that sends the necessary information to a BGP speaker in the parent domain. A separated policy can be configured to decide which information can be exported.

Note that the number of ”border links” that create the inter-domain network is usually quite low; even dynamicity is low: new links are seldom added (involving also commercial agreements between carriers) and connections crossing domain boundaries are less frequent than the intra-domain ones. Due to the implementation of different policies in the domains, routing information updates could be uncoordinated impacting on routing protocol convergence and leading to connection setup failures. Moreover, in the multi-carrier scenario, certain coordination about the kind of TE information to distribute may be required to avoid issues to the routing algorithms implemented at the inter-domain level. Consequently, another viable approach could be to configure the inter-domain links statically into the pPCE.

C. Hierarchical Stateful PCE

The IDEALIST project is considering the H-PCE [9], [12] as the most suitable technology to compute optimum routes for LSPs crossing multiple domains: The pPCE is responsible for domain sequence computation. Then, in each identified domain, a child PCE (cPCE) performs segment expansion.

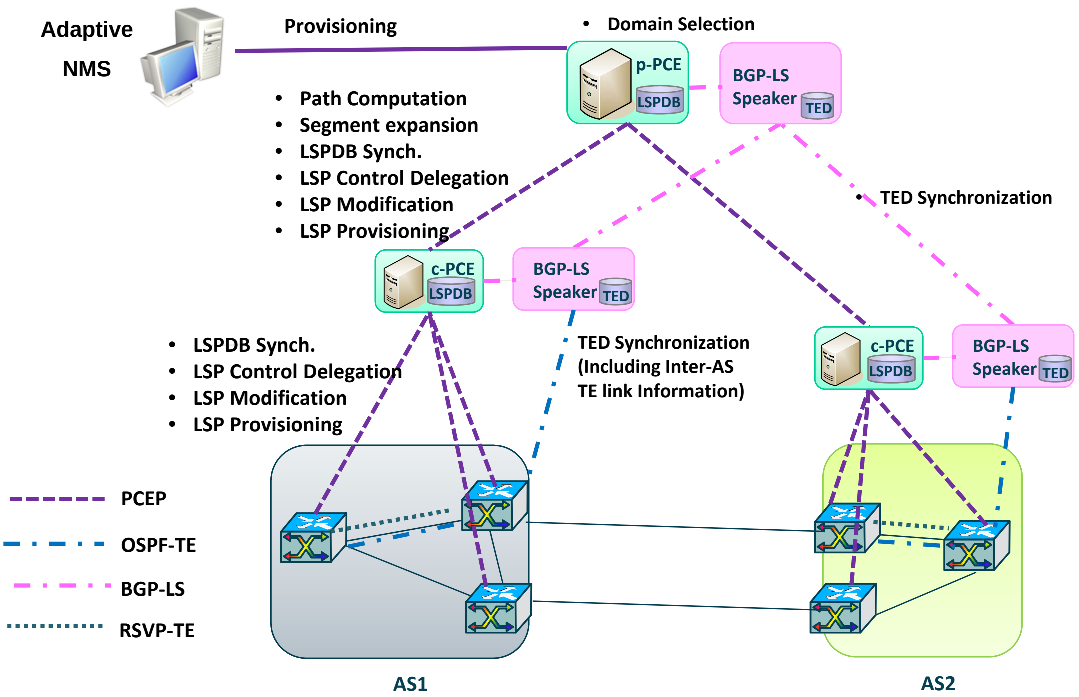

Fig. 2. Multi-domain control plane architecture based on a hierarchical stateful PCE and the use of BGP-LS as a northbound link state distribution protocol

The pPCE exploits an abstracted domain topology map that contains the child domains and their interconnections. In IDEALIST, several innovative enhancements to this approach are under investigation. First, besides reachability information, a mesh of abstracted links between border nodes is introduced in the parent TED to improve the effectiveness of domain sequence computation. Second, the north-bound distribution of Link-State and TE information using BGP (i.e., BGP-LS) is the protocol solution proposed to provide link information to the pPCE. Third, specific extensions to BGP-LS for elastic optical networks are also introduced [13]. Finally, a stateful condition is introduced at the pPCE to enable advanced TE solutions, e.g. multi-domain reoptimization. The approach completes the hierarchical path computation composed of domain sequence selection and segment expansion with a subsequent route segmentation and segment provisioning, as explained next.

D. Signalling Aspects

According to [14], inter-domain TE LSPs can be supported by one of three options: contiguous LSPs, stitched LSPs and nested LSPs. In the flex-grid context, the latter solution is not applicable. Since these solutions require a high degree of control plane interoperability both for routing and for signalling, we are considering, at the time being, taking advantage of the H-PCE structure, where the pPCE can orchestrate the cPCEs, acting as the responsible within its own domain, for the establishment (and release) of connections to an underlying GMPLS control plane. By this approach, all PCEs are stateful and have instantiation capabilities. That is, every domain has its own ”local” RSVP-TE session and the connectivity at the data plane level is insured by the concatenation of media channels at each domain, while the coordination among the domains (i.e. ingress/egress ports, labels, etc.) is the responsibility of the pPCE. In this case, interoperability requirements are scoped to PCEP extensions for stateful PCE with instantiation capabilities and no protocols are required at the inter-domain boundaries.

E. Control plane procedures

Let us detail the main procedure for the establishment of a LSP with the help of Figure 3. Each node, augmented with Path Computation Client capabilities, establishes a PCEP connection with the stateful PCE in its domain. A BGP-LS speaker (which could be co-located at the cPCE) is then responsible for distributing a policy-controlled abstract representation of the intra-domain connectivity towards the speaker that is located along the pPCE. Upon request (e.g. operator intervention), the ANM triggers the establishment of the connection by using the pPCE northbound interface (NBI). This NBI may, e.g., be based on a restricted PCEP subset, reusing the PCInitiate message format and related procedures. The pPCE then proceeds to perform a path computation that consists of a domain selection followed by the segment expansion [15]. Once the path Explicit Route has been obtained, it is segmented on a per domain basis and the corresponding segment is sent to each stateful cPCE using a PCInitiate message. This message is processed by the ingress node of the segment that continues with the subsequent RSVP-TE Path/Resv message exchange within the domain. The successful establishment is then reported first to the cPCE, which, in turn, reports to the pPCE. The latter finally composes the end result and sends it back to the ANM, with the notification of the successful multi-domain service provisioning. When the segments are established, the internal IGP protocols disseminate the changes which may trigger a change in the abstracted view of the domain, subject to policy.

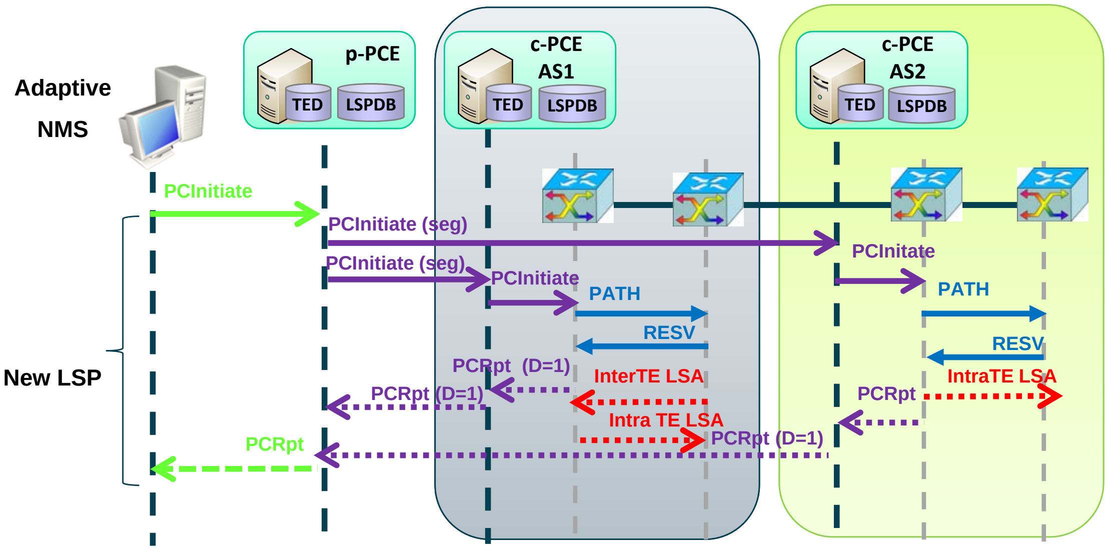

Fig. 3. Message flow detailing the establishment of a multi-domain elastic optical connection

IV. Conclusions

The design of a GMPLS/PCE control plane for flexi-grid optical networks presents different challenges, requiring architectural and protocol solutions that need to balance, on the one hand, the need for fulfilling the initial requirements and, on the other hand, design aspects such as robustness, security, and scalability. Although the GMPLS/PCE framework is considered to be stable and quite mature, having to address specific features regarding the optical technology and, in particular, with the constraints associated to flexi-grid DWDM networks, variable bandwidth transceivers and programmable devices is a complex problem. In this paper, we have detailed the major components of such a control plane. The current issues are related to the extension to the multi-domain context, characterized by topology visibility and administrative constraints. Whereas some aspects are well-known and understood, the summarization of TE capabilities per domain, underlay network abstraction and applicability of stateful PCE capabilities to end-to-end path computation across multi-domain networks is still a significant research problem. In this regard, IDEALIST proposes innovative solutions based on the hierarchization of the PCE and the use of dedicated protocols for topology dissemination.

Acknowledgment

This work was partially funded by the European Community's Seventh Framework Programme FP7/2007-2013 through the Integrated Project (IP) IDEALIST under grant agreement num. 317999.

References

[1] "ITU-T Recommendation G.872, Architecture of optical transport networks."

[2] "ITU-T Recommendation G.694.1, Spectral grids for WDM applications: DWDM frequency grid."

[3] Y. Lee, G. Bernstein, "Framework for gmpls and path computation element (pce) control of wavelength switched optical networks (wsons)," IETF RFC 6163, April 2011.

[4] E. Crabbe, J. Medved, and R. Varga, "PCEP extensions for stateful PCE, draft-crabbe-pce-stateful-pce-07," *IETF*, Oct. 2013.

[5] Zhang, F. and Zhang, X. and Farrel, A. and González-de-Dios, O. and Ceccarelli, D., "RSVP-TE Signaling Extensions in support of Flexible Grid," IETF I.-D. draft-zhang-ccamp-flexible-grid-rsvp-te-ext-03, work in progress, Nov 2013.

[6] A. Farrel, J. Drake, N. Bitar, G. Swallow and D. Ceccarelli, "Problem Statement and Architecture for Information Exchange Between Interconnected Traffic Engineered Networks," IETF draft-farrel-interconnected-te-info-exchange-03, work in progress, Feb 2014.

[7] Chen, M. and Zhang, R. and Duan, X., "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering," IETF RFC 5392, Jan 2009.

[8] D. King, A. Farrel, "The application of the path computation element architecture to the determination of a sequence of domains in mpls and gmpls," IETF RFC 6805, Nov 2012.

[9] R. Casellas *et al.*, "Dynamic virtual link mesh topology aggregation in multi-domain translucent WSON with hierarchical-PCE," in *ECOC*, Sept. 2011.

[10] F. Paolucci, O. Gonzalez de Dios, R. Casellas, S. Duhovnikov, P. Castoldi, R. Munoz, and R. Martinez, "Experimenting hierarchical pce architecture in a distributed multi-platform control plane testbed," in *Optical Fiber Communication Conference*, March 2012.

[11] Gredler, H. and Medved, J. and Previdi, S. and Farrel, A. and Ray, S., "North-Bound Distribution of Link-State and TE Information using BGP," IETF I.-D. draft-ietf-idr-ls-distribution-04, work in progress, Nov 2013.

[12] D. King and A. Farrel, "The application of the path computation element architecture to the determination of a sequence of domains in MPLS and GMPLS," *IETF, RFC6805*, Nov. 2012.

[13] M. Cuaresma *et al.*, "Experimental demonstration of H-PCE with BPG-LS in elastic optical networks," in *ECOC*, 2013.

[14] Farrel, A. and Vasseur, J.-P. and Ayyangar, A., "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering," IETF RFC 4726, Nov 2006.

[15] F. Zhang, R. Casellas, Q. Zhao, D. King, and O. Gonzalez de Dios, "Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)," IETF draft-ietf-pce-hierarchy-extensions (work in progress), February 2014.