# Utilization of Temporary Reservation of Path Computed Resources for Multi-Domain Path Computation Element Protocols in WDM Networks

Diego Álvarez*, Víctor López*†, José Luis Añamuro*, Jorge López de Vergara*,
Óscar González de Dios† and Javier Aracil*
*High Performance Computing and Networking group, Universidad Autónoma de Madrid, Spain. E-mail: victor.lopez@uam.es
†Telefónica I+D, Madrid, Spain, E-mail: ogondio@tid.es

***Abstract*—In recent years, Path Computation Element (PCE) architecture has been standardized as a suitable solution for path computation in multi-domain network scenarios. The Traffic Engineering Database (TED) of the PCE is updated with the information of the control plane. However, there is a delay from the PCE replies to a request, until the TED is updated with the network status information. This delay includes not only the control plane delay, but also Path Computation Element Protocol (PCEP) information exchange. This desynchronization in the TED and the real status information leads to an extra blocking situation when the Label Switch Router (LSR) tries to reserve a path, but it has been previously reserved by other LSR. To solve such problem, a recent draft is submitted to the IETF proposing new PCEP extensions for a pre-reservation of the computed path resources for a certain period.**

**This work implements in C three multi-domain algorithms: Per-domain Path Computation, Backward-Recursive PCE-based Computation and Hierarchical Path Computation Element and it assesses their performance with and without mechanisms to reduce this extra blocking probability due to the uncertainty of the TED information.**

## I. Introduction

With the advent of new Internet services, WDM networks are the solution to absorb such kind of high speed services that require low delay and high bandwidth utilization. The research community has done a great effort in the last years to provide a common control plane by standardizing Generalized Multi-Protocol Label Switching (GMPLS). GMPLS allows a dynamic and distributed configuration of the optical layer. However, the computation of optical paths becomes complex in terms of computation when the impairments induced by optical technologies are taken into account. If such computation is done into the GMPLS controller, the hardware requirements increase and, consequently, the node cost. In order to alleviate that issue, a Path Computation Element (PCE) architecture has been standardized [1].

In a PCE architecture, there is a PCE in each domain, which receives the request from the Path Computation Clients (PCC). As it is a single point of failure in the network, multiple PCEs can be located in the same domain [2]. PCE architecture fits with the requirements for multi-domain WDM scenarios [3]. The PCEs in a multi-domain scenario can cooperate as peers or in a hierarchical model [2]. There are three main PCE-based algorithms for multi-domain scenarios: Per-domain Path Computation [4], Backward-Recursive PCE-based Computation (BRPC) [5] and Hierarchical Path Computation Element (H-PCE) [6]. In the recent years, the research community has worked on the PCE architecture to improve the multi-domain path computation. Authors in [7] provide an overview of the developments in the area of PCE-based traffic engineering in GMPLS networks, analyze the BRPC approach in multi-domain networks in detail, and compare its performance with non-PCE existing solutions. Authors in [8] compare the performance of BRPC and per-domain algorithms, concluding that BRPC improves per-domain algorithm in terms of blocking probability. The behavior of the H-PCE is validated in [9] and the authors assess the computation time of the H-PCE but not its blocking probability. This work implements and compares these three multi-domain protocols in terms of blocking probability.

The PCE requires the network state information, which is stored in the Traffic Engineering Database (TED). This information is updated via OSPF Link-State Advertisement (LSA) messages [2]. When the TED information is different from the network state, the PCE can reply with resources that are already reserved. When the Label Switch Router (LSR) requests the path, the control plane denies the request since these resources are occupied. Let us call this type of block as "stolen-lambda" block. A recent draft is proposed in the IETF [10] to eliminate this kind of block using PCEP extensions to pre-reserve resources. To the best of the authors knowledge, this is the first work that implements such extensions and we show the avoidance of the "stolen-lambda" block thanks to this mechanism. A PCE Proactive scheme is proposed in [11], which is similar to IEFT draft [10]. The PCE Proactive scheme does not support timers for the reservation like [10] and its validation in [11] is just done by simulation.

The remaining is organized as follows: Section II provides the basics of the multi-domain PCE algorithms. Section III defines the impairment-aware routing used in this work. The evaluation between the algorithms with the implemented extensions is presented in section IV. Finally, section V concludes the paper.

## II. Multi-domain Path Computation Element protocols

The 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" [1]. Path Computation Element Protocol (PCEP) follows a request/response scheme, where a PCC asks for routes to the PCE. The PCE is in charge of carrying out the route computation by taking into account the physical restrictions and the available resources in the optical layer. The PCE is aware of the network state information thanks to the flooding information mechanisms used in GMPLS. For networks without a control plane, the PCE can be located as part of the management plane [2]. When client equipment requests a new path, the request is sent via the user-to-network interface (UNI) as in traditional GMPLS networks. However, the GMPLS controller of the optical equipment redirects the request to the PCC.

One of the main motivations behind PCEs deployment is to tackle the problem of multi-domain Label Switched Paths (LSPs) establishment. There are three different computation methods namely Per-domain Path Computation, Backward-Recursive PCE-based Computation (BRPC) and Hierarchical PCE (H-PCE). Each method is discussed in the following.

*1) Per-domain Path Computation:* At this approach, the path is computed during the signalling process domain by domain. Each PCE computes the path from its ingress to egress router in its domain [4]. Consequently, the sequence of domains to be traversed must be known beforehand by the PCE in the source domain. However, there is not a mechanism to choose the best domains from the source to the destination. In addition, this procedure provides suboptimal paths because if there are multiple connections between the domains, the PCE may provide a path that is optimal locally, but not overall.

*2) Backward-Recursive PCE-based Computation:* This procedure is based on comunication and cooperation between PCEs to compute optimal interdomain paths. The BRPC method starts at the destination domain, which sends to its neighbor a tree of potential paths from every ingress node to the destination node. Each PCE in the domain sequence adds its own paths from its ingress nodes to the tree and passes it to the previous domain. This process continues until the source domain is reached, which selects the best end-to-end path. Fig. 1 depicts three connected domains with one PCE per domain. Using BRPC, the PCE1 sends a request to the PCE2, which forwards it to the PCE3. The PCE3 replies with the distance from its ingress nodes (N and O) with the domain 2. The PCE2 carries out the same operation sending a tree with the possible combinations between the edge nodes from domain $1$ to $3$. When multiple domains are interconnected such information exchange can be complicated. If the sequence of domains is known, this process is easier. Border Gateway Protocol (BGP) can be used for this purpose. Unfortunately, BRPC does not scale with complex multi-domain topologies.

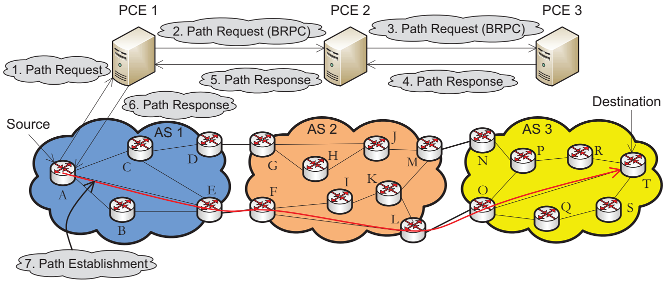

Fig. 1. Backward recursive path computation

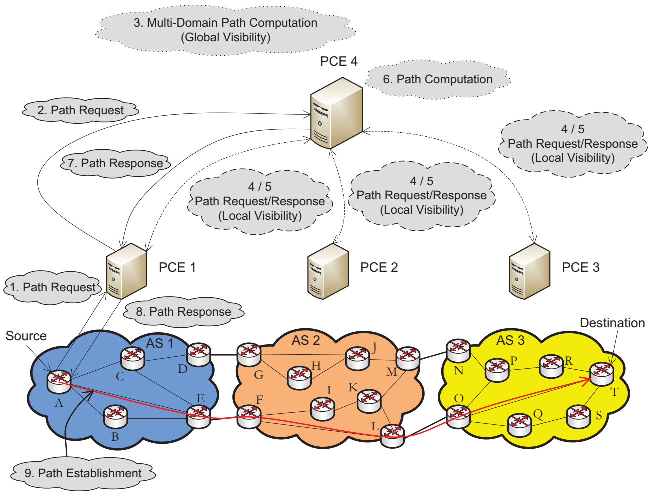

Fig. 2. H-PCE multi-domain path computation

*3) Hierarchical PCE:* Fig. 2 shows an example of a H-PCE architecture. In this architecture there is a parent PCE and some child PCEs, and they are organized in multiple levels [6]. The parent PCE does not have information of the whole network, but is only aware of the connectivity among the domains and provide coordination to the child PCEs. The path request is sent to the parent PCE, which selects a set of candidate domain paths and sends requests to the child PCEs responsible for these domains. Then the parent PCE selects the best solution and it is transmitted to the source PCE. This hierarchical model fits with the model for the Automatic Switched Optical Network (ASON), since the networks are composed by sub-networks and the routing areas have relationship between peers.

## III. RWA-algorithm with impairment awareness

The following well-known algorithms are used in the Routing and Wavelength Assignment (RWA) units, namely Dijkstra's $k$-shortest path and first-fit (FF). Both are very simple in terms of their implementation. The cost used in the routing problem may be set to distance, number of hops, link load, etc. such that depending on the scenario a route is determined based on a particular cost. The FF algorithm chooses the first available wavelength until there are no more resources available.

The characteristics of the optical network elements such as fiber and nodes should be made available to the entity that makes the IA-RWA decisions (here: the PCE). The most important linear impairments are optical signal to noise ratio (OSNR), residual chromatic dispersion (CD), polarization mode dispersion (PMD) and, in case of lightpaths, hop and technology-dependent penalties due to filter cascading at intermediate nodes [12]. The latter also depends on whether or not walk-off is present in the transmitter's wavelength. Furthermore, the optical fiber infers several non-linear impairments such as phase noise, self-phase and cross-phase modulation, and four-wave mixing. These are not always noticeable because the effects are highly depended on the used bit rate and modulation format.

To the best of the authors' knowledge, commercially available routers or sub-modules may only provide real-time data on the wavelength, and received power and OSNR of an active optical channel. Even though many research papers exist on the real-time measurement of several impairments, considering cumulated impacts is widely accepted by means of applying margins on the required OSNR and/or on the experienced non-linear phase shift. In this work, the expected OSNR and PMD values are considered to be of key importance while all other effects are covered by a single OSNR penalty in dB.

The ITU-T standard G.680 provides a detailed equation to calculate the received OSNR of a lightpath that takes into account individual properties of each span per wavelength [13]. A well-known simplified version is as follows

$$OSNR = P_{\text{out}} - \alpha L - 10 \log_{10} N + 58 - F, \tag{1}$$

with the OSNR in dB, $P_{\text{out}}$ as the signal power in dBm at the output of the last amplifier, $\alpha$ as the attenuation factor in dB/km, $L$ as the length of the fiber span in km, $N$ as the total number of spans, and $F$ the optical amplifier's noise factor in dB. The value of 58 in (1) equals to $10 \log_{10}(h f_i \Delta f_0)$ with $h$ as Planck's constant, $f_i$ as the optical frequency in Hz, and $\Delta f_0$ as the reference bandwidth in Hz. Accordingly, the PCE evaluates the following OSNR-condition for a path coming from the Routing-unit

$$OSNR > OSNR_{\text{min}} + OSNR_{\text{imp}}, \tag{2}$$

with $OSNR_{\text{min}}$ as the minimum required OSNR-value for error-free detection and $OSNR_{\text{imp}}$ as the cumulated penalty for other (non)-linear impairments. All parameters in (1) are known so the maximum number of spans $N_{\text{max}}$ fulfilling (2) can be determined.

Regarding the experienced PMD, the equation in [13] includes the PMD values of all optical network elements constituting the lightpath. However, [14] shows a simplified version that is used to check the physical constraints namely

$$PMD = B \sqrt{\sum_{k=1}^{N} D_{\text{PMD}}^2 \cdot L}, \tag{3}$$

with $B$ as the channel symbol rate, and $D_{\text{PMD}}$ as the PMD value of each fiber span. Typically, less than 10% pulse broadening ($a = 0.1$) is recommended for a lightpath, and therefore (3) is evaluated on being smaller or equal than $a$. Similar to the OSNR-limit, the value for $N_{\text{max}}$ can also be determined for the PMD-limit and the smallest number should be stored.

It is clear that several assumptions have been made to justify the simplifications applied in (1) and (3). Regarding the optical amplifiers, the gain $G$ fully compensates the transmission losses of a span ($G = \alpha L$), gain-control is in place such that the gain received by an individual channel agrees with the former, and all amplifiers have an equal noise factor. Regarding the fiber spans, it is assumed that these are all of equal length with average values for the attenuation factor and PMD value. Regarding the factor $10 \log_{10}(h f_i \Delta f_0)$, the difference between wavelengths on the blue or the red-side of a wavelength comb is negligible such that the value 58 in (1) is valid for all channels.

To sum up, the only lightpath-dependent parameter in (1) and (3) is $N$ which reduces the process to check the OSNR and PMD limits to determining $N$ for a path and to compare this value with $N_{\text{max}}$. Future PCE implementations may consider a more exact treatment of the physical impairments.

## IV. Experimental Results

We have run the experiments using the Chinese core network [15], which is shown in Fig. 3. For this study the maximum number of wavelengths is set to $M = 80$ and $K = 5$ in the RWA algorithm. Regarding the physical impairments, $P_{out} = 4$ dBm, $\alpha = 0.35$ dB/Km, $L = 80$ km, $OSNR_{imp} = 1.5$ dB, $OSNR_{min} = 12$ dB, $B = 10$ Gbps and $D_{PMD} = 0.2\ 10^{-12}$ s/km$^{0.5}$. An $OSNR_{imp}$ margin is taken as the cumulated penalty of all other impairments which is similar to OSNR margins found in data sheets of commercial 10-Gigabit transceivers.

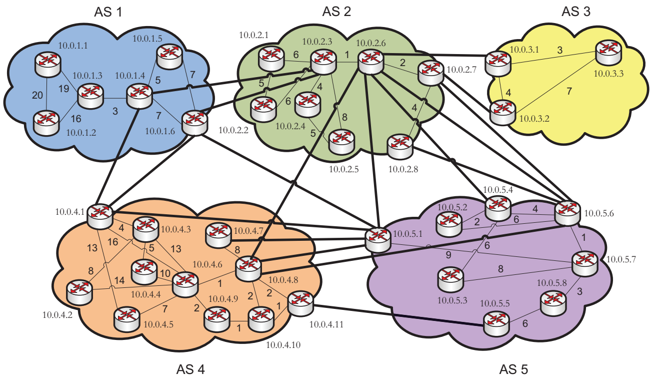

Fig. 3. Backbone Chinese topology [15]

For the experimental results, we have defined a PCC for each domain, which is requesting the routes to the PCE in its domain. For this work, we have implemented in C the PCEP protocol, the extensions to support the three multi-domain methods as well as the pre-reservation mechanism. Call requests for each node pair follow a Poisson process for the arrival rate and the holding time. We have run 20000 requests to achieve stable results. The PCEs and PCCs are in the same server (4-core Intel(R) Xeon(R) CPU E5345 at 2.33 GHz and 8GB of RAM memory). In this light, the delays in our experiment are the PCEP protocol and the RWA algorithm with physical impairments defined in section III. There is a propagation delay (the PCEs are located in different places), which is not included in our study. The TED information is updated by the PCC when they receive the response. Depending on the PCE location, the number of control plane hops and the updating timer for the LSA messages, this time can be important and it adds an uncertainty to the path computation. This effect is out of the scope of this work.

Regarding the sequence of domains for the route computation, the BRPC and H-PCE use K-shortest path. $K$ is higher for H-PCE case because it has a more detailed information about the connections between domains than in BRPC.

### A. Impact of "stolen-lambda" block

Fig. 4 shows the blocking probability of the BRPC algorithm ($K = 4$) and the reason of the block: (1) there is no wavelength available, (2) the physical restrictions are not fulfilled and (3) the "stolen-lambda" block. The "stolen-lambda" block appears when the PCE is replying to two requests with a route which shares resources. The first request reserves the resources, but when the second request tries to reserve them they are no longer available. In this work, the TED is updated by the PCCs when they receive the response. In a real network, the propagation and the control plane delay increase the differences between the network state and the PCE TED, thus incrementing the blocking probability.

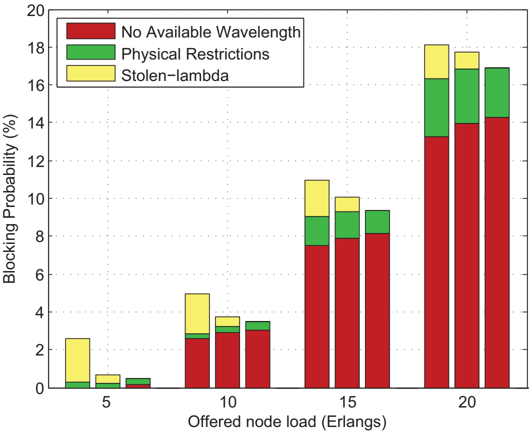

Fig. 4. Blocking probability of the BRPC with First-Fit (left), Round-Robin First-Fit (middle) and Pre-Reservation (right) algorithms

When using the RWA algorithm with FF (Fig. 4 left-bars), there is an important contribution of the "stolen-lambda" effect. To reduce this "stolen-lambda" block, we used a modification of First-Fit called Round-Robin First-Fit (RR-FF). As the PCE is using the FF mechanism, it starts the assignment process from lambda 0. Instead of starting always from lambda 0, if the PCE responses with a lambda $n$, RR-FF begins the search from lambda $n+1$ in the next request. Fig. 4 (middle-bars) shows the blocking probability of the RWA with RR-FF. This mechanism is simple, it does not add complexity to the algorithm and it reduces the blocking probability due to the "stolen-lambda" effect, but not eliminates it. Pre-Reservation (PR) algorithm results are discussed in the next section (Fig. 4 right-bars).

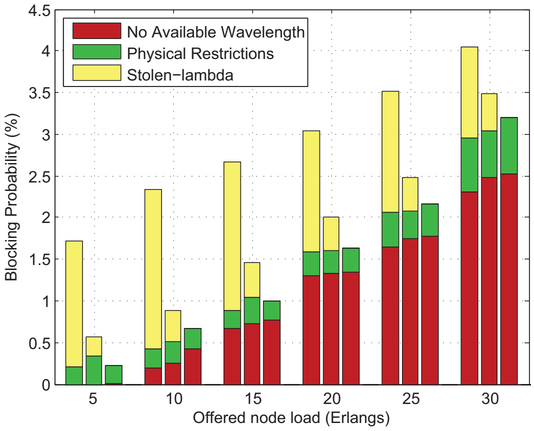

Fig. 5. Blocking probability of the H-PCE with First-Fit (left), Round-Robin First-Fit (middle) and Pre-Reservation (right) algorithms

H-PCE achieves a lower blocking probability than BRPC when using $K = 20$ (Fig. 5). The time consumption per request is higher in the H-PCE than in the BRPC. Fig. **??** shows the average delay per request for different $K$ values using FF. The computation algorithm in the H-PCE is carried out by the parent PCE and it sends requests to the child PCEs sequentially to obtain the path segments. Then, it concatenates them and calculates the optimal path. This process implies a higher time consumption as it increases $K$, but the results in terms of blocking probability are better. The "stolen-lambda" block is more important in the H-PCE case than in the BRPC case (Fig. 4) in relation with the total blocking probability. For lower values of $K$, the time per request is very small thus reducing the "stolen-lambda" block or even eliminated at low load scenarios for BRPC and H-PCE.

### B. Performance with temporary reservation of computed resources

Authors in [10] propose a temporary reservation of the resources to avoid the "stolen-lambda" block. This mechanism pre-reserves the resources in the PCE TED when there is a path request for a given time ($T_{res}$). Once the $T_{res}$ timer expires, the PCE removes the reservation state of such resources. If the path was established, the OSPF LSA messages update the PCE TED properly. When using PR mechanism, the "stolen-lambda" block is eliminated if $T_{res}$ is long enough ($T_{res} = 3s$ in our experiments). Blocking probability of PR mechanism is shown in Fig. 4 and Fig. 5 (right-bars). There is an important reduction of the blocking probability, when using the PR mechanism. The "stolen-lambda" block is completely

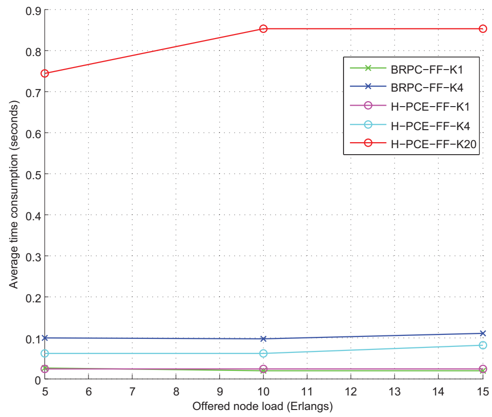

Fig. 6. BRPC and H-PCE average time consumption per request using FF

eliminated. There is a slight increment of the block because there are no more available wavelengths. The reason is that as the requests are not rejected because of the "stolen-lambda" effect, they occupy resources thus incrementing the blocking probability. Let us remark that if as there are differences from the network state information and the TED, the timer $T_{res}$ should be incremented to obtain similar results.

Fig. 7 shows the blocking probability for per-domain, BRPC and H-PCE methods using temporary reservation. In light of this, we can conclude that H-PCE reduces the blocking probability in multi-domain scenarios and, thanks to the temporary reservation the "stolen-lambda" effect can be eliminated. For $K = 1$, BRPC and H-PCE only check a multi-domain path, thus achieving results similar to per-domain method.

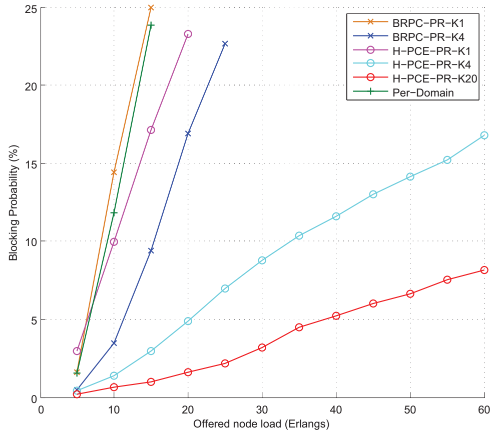

Fig. 7. Comparison in terms of blocking probability using temporal reservation

## V. Summary and Conclusions

The contribution of this work is two-fold: (1) the comparison of per-domain, BRCP and H-PCE in terms of blocking probability and (2) the validation of temporary reservation mechanism as a solution to avoid "stolen-lambda" block. H-PCE has a better performance in terms of blocking probability and its scalability is better than BRPC mechanism.

As future work, we will evaluate the algorithms behavior including the control plane uncertainty and the propagation delay between the PCEs and PCCs.

## Acknowledgements

This work has been funded by the Spanish Ministry of Education and Science under project ANFORA (TEC2009-13385).

## References

[1] A. Farrel, J.P. Vasseur, and J. Ash, "A path computation element (PCE)-based architecture," *IETF RFC 4655*, pp. 1–40, Aug. 2006. Online (Nov. 2009): http://tools.ietf.org/html/rfc4655.

[2] V. López, B. Huiszoon, J. Fernández-Palacios, O. González de Dios, and J. Aracil, "Path computation element in telecom networks: Recent developements and standardization activities," in *Proc. Optical Networking Design and Modeling (ONDM)*, pp. 1-6, Feb. 2010, Kyoto, Japan.

[3] M. Chamania and A. Jukan, "A survey of inter-domain peering and provisioning solutions for the next generation optical networks," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 1, pp. 33–51, 2009.

[4] J.P. Vasseur (Ed.), A. Ayyangar (Ed.) and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)," *IETF RFC 5152*, pp. 1–21, February 2008. Online: http://tools.ietf.org/html/rfc5152.

[5] J.P. Vasseur (Ed.), R. Zhang, N. Bitar and J.L. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths," *IETF RFC 5441*, pp. 1–18, April 2009. Online: http://tools.ietf.org/html/rfc5441.

[6] F. Zhang, Q. Zhao, O. Gonzalez de Dios, R. Casellas and D. King, "Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)," *IETF Internet-Draft*, pp. 1–14, April 2011. Online: http://tools.ietf.org/html/draft-zhang-pce-hierarchy-extensions-00.

[7] S. Dasgupta, J.C. de Oliveira, and J.P. Vasseur, "Path-computation-element-based architecture for interdomain MPLS/GMPLS traffic engineering: Overview and performance," *IEEE Network*, vol. 21, no. 4, pp. 38–45, July 2007.

[8] R. Casellas, R. Martínez, R. Muñoz, and S. Gunreben, "Enhanced Backwards Recursive Path Computation for Multi-area Wavelength Switched Optical Networks Under Wavelength Continuity Constraint," *Journal of Optical Communications and Networking*, vol. 1, no. 2, pp. A180–A193, 2009.

[9] R. Casellas, R. Muñoz, and R. Martínez, "Lab trial of multi-domain path computation in gmpls controlled wson using a hierarchical pce," in *Optical Fiber Communication Conference*. Optical Society of America, 2011.

[10] O. Gonzalez de Dios, R. Casellas and F.J. Jimenez, "PCEP Extensions for Temporary Reservation of Computed Path Resources and Support for Limited Context State in PCE," *IETF RFC draft*, pp. 1–15, Oct 2010. Work in progress (Oct. 2010): http://tools.ietf.org/html/draft-gonzalezdedios-pce-reservation-state-00.

[11] A. Giorgetti, F. Cugini, N. Sambo, F. Paolucci, N. Andriolli, and P. Castoldi, "Path state-based update of PCE traffic engineering database in wavelength switched optical networks," *Communications Letters, IEEE*, vol. 14, no. 6, pp. 575 –577, June 2010.

[12] A.H. Gnauck, P.J. Winzer, C. Dorrer, and S. Chandrasekhar, "Linear and nonlinear performance of 42.7-Gb/s single-polarization RZ-DQPSK format," *IEEE Photon. Technol. Lett.*, vol. 18, no. 7, pp. 883–885, April 1 2006.

[13] ITU-T, "Physical transfer functions of optical network elements," *G.680 Recommendation*, pp. 1–68, Jul. 2007. Online (Nov. 2009): http://www.itu.int/rec/T-REC-G.680/en.

[14] J. Strand, A.L. Chiu, and R. Tkach, "Issues for routing in the optical layer," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 81–87, Feb. 2001.

[15] Y. Zhao, J. Zhang, Y. Ji, and W. Gu, "Routing and Wavelength Assignment Problem in PCE-Based Wavelength-Switched Optical Networks," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 2, no. 4, pp. 196–205, 2010.