A Survey on the Path Computation Element (PCE) Architecture

Francesco Paolucci, Filippo Cugini, Alessio Giorgetti, Nicola Sambo, and Piero Castoldi

***Abstract*—Quality of Service-enabled applications and services rely on Traffic Engineering-based (TE) Label Switched Paths (LSP) established in core networks and controlled by the GMPLS control plane. Path computation process is crucial to achieve the desired TE objective. Its actual effectiveness depends on a number of factors. Mechanisms utilized to update topology and TE information, as well as the latency between path computation and resource reservation, which is typically distributed, may affect path computation efficiency. Moreover, TE visibility is limited in many network scenarios, such as multi-layer, multi-domain and multi-carrier networks, and it may negatively impact resource utilization.**

The Internet Engineering Task Force (IETF) has promoted the Path Computation Element (PCE) architecture, proposing a dedicated network entity devoted to path computation process. The PCE represents a flexible instrument to overcome visibility and distributed provisioning inefficiencies. Communications between path computation clients (PCC) and PCEs, realized through the PCE Protocol (PCEP), also enable inter-PCE communications offering an attractive way to perform TE-based path computation among cooperating PCEs in multi-layer/domain scenarios, while preserving scalability and confidentiality.

This survey presents the state-of-the-art on the PCE architecture for GMPLS-controlled networks carried out by research and standardization community. In this work, packet (i.e., MPLS-TE and MPLS-TP) and wavelength/spectrum (i.e., WSON and SSON) switching capabilities are the considered technological platforms, in which the PCE is shown to achieve a number of evident benefits.

***Index Terms*—GMPLS, PCE, PCEP, Path computation, routing, multi-domain, multi-layer, MPLS, WSON, SSON.**

I. Introduction

THE INCREASING demand for applications and services demanding for flexible and guaranteed Quality of Service (QoS) has pushed network operators to adopt MultiProtocol Label Switching (MPLS) and Generalized MPLS (GMPLS) control plane in core networks [1]. (G)MPLS provides the Traffic Engineering (TE) capability to route traffic flows, namely Label Switched Paths (LSPs), along explicit routes. Thanks to resource availability and topology information collected through routing protocols (e.g., Open Shortest Path First with TE extensions, OSPF-TE), such TE capability allows source nodes to perform path computation subject to additional QoS constraints typical of such networks, e.g., guaranteed bandwidth in MPLS networks, wavelength continuity constraint in Wavelength Switched Optical Networks (WSONs).

Manuscript received May 3, 2012; revised October 1, 2012.

F. Paolucci, A. Giorgetti, N. Sambo and P. Castoldi are with TeCIP Institute, Scuola Superiore Sant'Anna, Via Moruzzi 1, 56124 Pisa, Italy (e-mail: fr.paolucci@sssup.it).

F. Cugini is with Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Via Moruzzi 1, 56124 Pisa, Italy.

Digital Object Identifier 10.1109/SURV.2013.011413.00087

In LSP provisioning, the path computation process represents one of the crucial steps to achieve TE solutions and an adequate network resource utilization. In single domain networks the path computation is usually determined by the routing process at the source node, while resources are reserved during the signaling phase, exploited through distributed protocols, such as the Reservation Protocol with TE extensions (RSVP-TE). In dynamic network scenarios the separation between the two operations may lead to sub-optimal TE solutions generally inducing waste of network resources. In optical networks, as for example in WSONs, the wavelength assignment process, typically performed by the destination node, may introduce additional potential TE inefficiencies. Moreover, distributed path computation may require heavy processing at each control plane node, especially when based on multiple constraints.

Moving from single-domain single-layer scenario to multi-layer, multi-domain networks, additional issues arise, such as restricted topology visibility due to scalability reasons. Moreover, when multiple carriers are involved in a path computation, i.e. source and destination of a traffic request belong to different administrative domains, the need to preserve information confidentiality across domains prevents the open advertisement of detailed intra-domain network resources. Such limitations considerably complicate path computation and affect the inter-layer/inter-domain TE performance in terms of the overall network resource utilization. As a matter of fact, network operators do not currently implement inter-domain TE techniques and the provisioning of QoS-guaranteed applications across multiple domains is performed manually, often requiring several weeks, and it typically relies on sub-optimal solutions (i.e., intra-domain independent path computations).

The aforementioned path computation limitations for provisioning operations are at the basis of a significant research activity carried on in the last years and still active nowadays in the context of core networks control plane.

The Internet Engineering Task Force (IETF) has proposed a set of techniques defined under the umbrella of the Path Computation Element (PCE) Architecture [2]. Such techniques rely on path computation performed by dedicated network entities (i.e., the PCEs).

Considerable research activity has been focused on the PCE architecture in the last years. In particular, several research projects involving both academia and important industrial partners have been funded by the European Commission (e.g., GEYSERS, ETICS, STRONGEST projects [3]–[5]) and the National Science Foundation (e.g., the DRAGON project [6]).

The PCE collects link-state information and performs path computation on behalf of network nodes. In addition, it may resort to other information sources, such as the network management system (NMS), to retrieve detailed information about resource utilization (e.g., used wavelengths) or physical network parameters (e.g. link length, impairments). The PCE provides the additional advantage that network nodes can avoid highly CPU-intensive path computations and effective TE solutions are achievable also in case of legacy network nodes.

Communication between a network element, referred to as Path Computation Client (PCC) and the PCE is achieved by exploiting the Path Computation Element communication Protocol (PCEP) [7]. The PCE has the responsibility of the path computation in its own layer/domain, where it has full visibility and updated information on available network resources. Cooperation between PCEs takes place in multi-layer/domain scenarios by sharing the result of each (intra-domain) path computation expressed as, for example, border node(s) to traverse, encrypted intra-domain routes, metric values. The combination of these results provides the entire source-destination path, and no additional information is exchanged among different domains. Specific techniques and procedures have been proposed and defined for each specific scenario.

This survey aims at providing the main activity efforts in utilizing the PCE architecture in the context of single-domain, multi-layer and multi-domain GMPLS networks, with particular focus dedicated to packet-switched MPLS networks, lambda-switched (i.e., WSON) and the recently introduced spectrum switched optical networks (SSON). The role of PCE is analyzed and discussed for what concerns LSP provisioning and also reliability aspects. In Sec. II we first provide a general overview of the PCE Architecture, comprising the main motivations for the PCE adoption. In Sec. III the main PCEP protocol messages, objects and operations are detailed. Then, we analyze the PCE applicability in various network scenarios, starting from the simplest single domain single-layer network, described in Sec. IV and going through increasing complexity scenarios, such as multi-layer networks (MPLS-based packet-switched layer and WSON-based lambda-switched layer), described in Sec. V, multi-domain networks in Sec. VI and multi-domain multi-carrier scenarios in Sec. VII. Finally, conclusions are drawn in Sec. VIII.

II. PCE ARCHITECTURE

The PCE architecture relies on two functional elements: the PCE and the PCC. The PCE, possibly implemented on a dedicated server, is responsible to perform constraint-based path computation requested by PCCs. A PCC is typically implemented on a Network Management System (NMS) or a network node (e.g., the requesting LSP source node). A PCE may also behave as PCC requesting path computations to other (peer or hierarchically higher) PCEs. PCC and PCE communicates through PCEP [7], described in Sec. III.

The typical PCE internal architecture is depicted in Fig. 1. The main component modules are the *Traffic Engineering Database* (TED), the *path computation module* and the *communication module*. The TED collects information about network topology and current TE information (e.g., link bandwidth utilization), populated by routing protocols session (e.g., OSPF-TE) or other mechanisms (e.g., by the NMS). The strategies adopted to create and keep updated the PCE TED are discussed in Sec IV-C. The path computation module is responsible for the computation of requested paths based on selected algorithms and specific policies. The communication module acts as interface to handle communication protocols (e.g., routing protocols and PCEP).

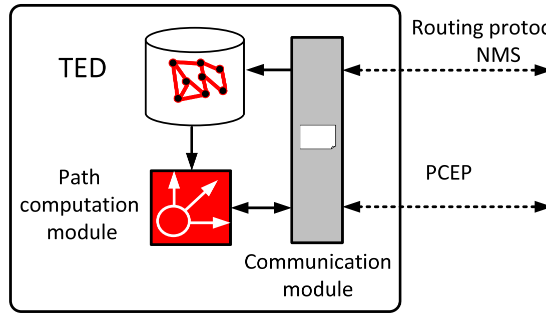


Fig. 1. Functional modules of a Path Computation Element.

The PCE can be either co-located within a network node (i.e., *internal PCE*) or a dedicated physical device (i.e., *external PCE*). Without loosing generality, this paper typically considers external PCEs.

The computation of an LSP can be performed resorting either to a *single PCE* or *multiple PCEs*. In the former case, depicted in Fig. 2-a the PCE returns the detailed end-to-end path route (step 1-2, dotted lines), whereas in the latter, depicted in Fig. 2-b and Fig. 2-c, more than one PCEs are involved in path computation, either in independent, peer or hierarchical fashion. In particular, in multiple PCE computation, each PCE computes a path segment (e.g., in Fig. 2-b PCE1 computes A-B-C at step 1-2, PCE2 computes C-D at step 5-6), in multiple inter-PCE computation, end-to-end path is the result of collaborative computation among PCEs (e.g., in Fig.2-c upon request reception at step 1, PCE1 forwards the request to PCE2–step 2–which provides C-D segment at step 3, then PCE1 computes the segment A-B-C, stitches the two segments and provides the end-to-end path to source node at step 4).

The temporal relationship between path computation and signaling depends on the considered network scenario and the selected PCE architecture. In a single PCE scenario, first the source node (acting as PCC) requests end-to-end path computation to PCE, then it triggers signaling protocol to reserve resources along the computed path. The process is sketched in Fig. 2-a. In the case of multiple PCEs, two alternatives are considered: 1) signaling is triggered upon each segment path computation (see Fig. 2-b), 2) signaling is triggered upon end-to-end inter-PCE path computation (see Fig. 2-c).

The general PCE architecture requires that each PCC is aware of the presence and the location of a PCE in the controlled domain and the definition of its path computation area. PCE discovery mechanisms have been proposed in [8]. They implement automatic and dynamic detection of PCEs

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

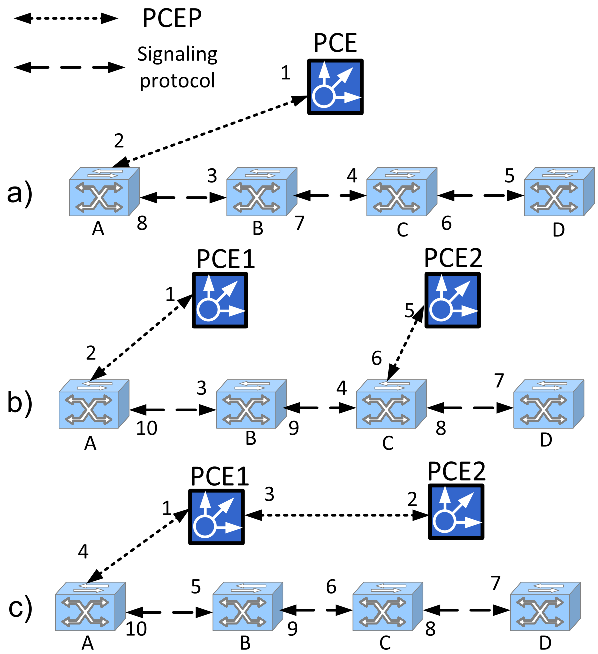


Fig. 2. Single PCE computation (a), multiple PCE computation (b), multiple inter-PCE computation (c).

along with additional information about supported capabilities and target area/domain, in order to perform the most suitable PCE selection for a given path computation request.

The PCE performs path computation based on its internal Traffic Engineering Database (TED), that contains the network topology and the current TE information (e.g., link bandwidth utilization). Based on the considered network scenario, a PCE is responsible for path computation on a specific network area. To this extent, the PCE TED visibility is restricted to a single area, domain or layer and the PCE performs path computation requests having, as source, a node belonging to the TED. Therefore, a single path computation request is generally sufficient to provision intra-area/domain/layer LSP, while path computation of LSPs traversing more areas/domains/layers requires a coordination among each involved PCE.

This survey describes the adoption of the PCE architecture in several network scenarios (e.g., single domain, multi-layer, multi-domain, multi-carrier) as proposed in the literature, in research projects and in the standardization bodies in the last years. The tree depicted in Fig. 3 illustrates the main research issues in the context of the PCE architecture, along with related works considered in the survey. The tree also reflects the survey structure starting from Sec. IV.

A. PCE motivations

The work in [9] provides a detailed overview of distributed TE solutions applied in legacy IP networks (e.g., IP over ATM), and extendible to IP over WDM networks using (G)MPLS. Similar TE solutions can also be applied in emerging network scenarios such as carrier grade Ethernet and MPLS-TP. However, a list of motivations is more recently provided in [2] for the adoption of a centralized PCE architecture.

First, intensive path computation algorithms requiring high CPU resources can be performed by a dedicated network element. The placement of a bundle of LSPs aiming at optimizing a given performance metric, or multi-constraint path computation including link resource utilization are typical examples of intensive algorithms. The second reason relies on the restricted topology visibility provided by routing protocols (e.g., Interior Gateway Protocols (IGP) family, such as OSPF-TE). Topology restriction may occur in case of scalability issues (e.g., multi-area and multi-domain networks), switching capability diversity (e.g., multi-layer), confidentiality and business critical issues (e.g., in multi-carrier networks). Cooperating PCEs placement overcomes such limitations allowing TE solutions in the aforementioned scenarios. Another reason is that some network devices do not run a control plane (e.g., legacy optical nodes equipment). In addition, the PCE makes synchronized path computation feasible. Such possibility, described in Sec. IV-A, represents a valuable advantage for many services provided by network operators requiring a set of LSPs with specific combined constraints (e.g., QoS-based virtual private networks, VPN) or requiring joint optimization. Finally, PCE has the capability to perform diverse path computation (e.g., primary and backup path, or fast reroute detours).

III. PCE COMMUNICATION PROTOCOL (PCEP)

The first IETF draft describing PCEP was published in November 2005. The main protocol structure is specified in [7], and consists of a client-server interaction between PCC and PCE. The PCC can be a network node (e.g., ingress node), a network operator, the NMS, or another PCE. Interaction is achieved through the exchange of PCEP messages running over TCP/IP, to exploit its reliability. Messages are defined to initiate, maintain and terminate a PCEP session.

To perform path computations, PCC and PCE first open a PCEP session within a TCP session. The PCEP session establishment includes the exchange of `Open` and `Keepalive` messages in order to agree on session parameters, such as timers and session refresh messages.

The core of the protocol interaction is realized through the exchange of two PCEP messages: the `Path Computation Request` (`PCReq`) message and the `Path Computation Reply` (`PCRep`) message. Additional messages are also defined to handle specific events and communication errors (e.g., `Error` (`PCErr`) and `Notification` (`PCNtf`) messages). The PCEP session terminates upon the reception of a `Close` message. A number of PCEP objects have been proposed and defined. The main path computation parameters defined in [7] and in additional RFCs within the IETF PCE Working Group are summarized, along with the related PCEP objects, in Table I.

A. PCEP Path Computation Request Message

A path computation request is included within a `PCReq` message specifying all the requested parameters and constraints. The `PCReq` message requires two mandatory objects:

- `End-Points Object`: contains the IP addresses of the source and destination of the required path, which may correspond to the end nodes of an entire path (i.e., LSP) or of a path segment.

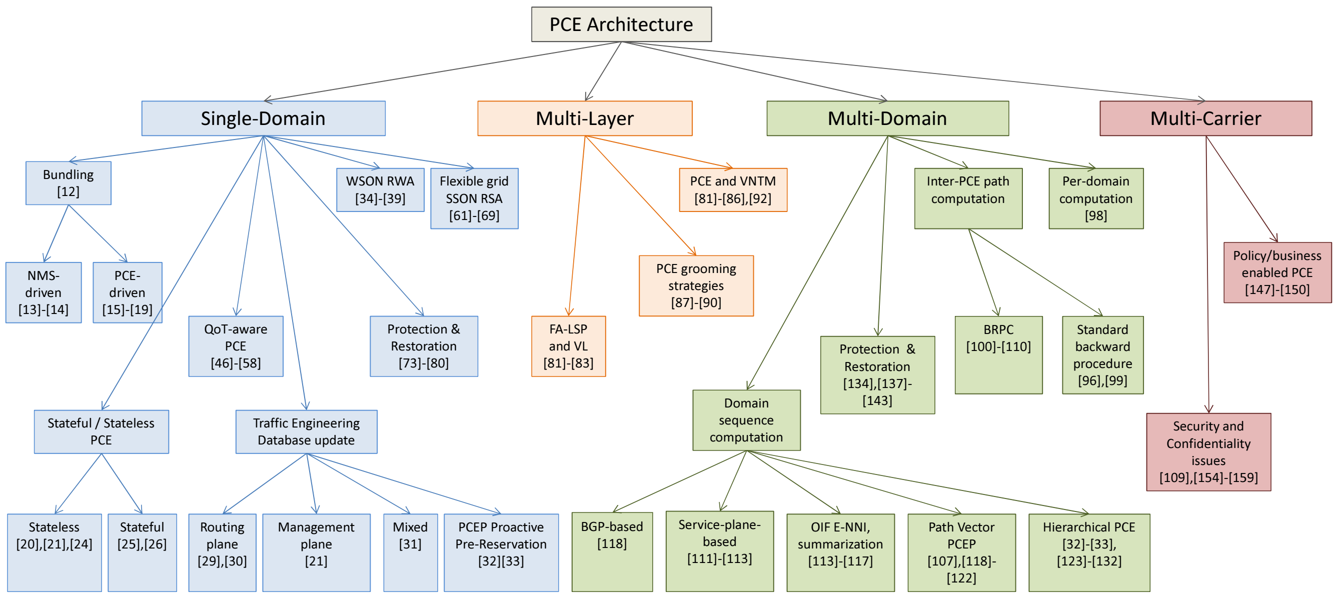


Fig. 3. Survey tree structure: scenarios and related solutions adopting the PCE architecture proposed in the literature.

- `Requested Parameter Object` (`RP`): encloses the request-ID used to uniquely identify each path computation request triggered by a PCC. Flags are specified to require computation priority, re-optimization of a previously established path, bidirectional path computation (with the same TE parameters in both directions) and whether strict or loose routes are expected. In strict routes, the list of nodes to be traversed is explicitly indicated (i.e., strict nodes) while in loose routes it is represented by a partial sequence or abstract nodes.

Additional objects may be included in the `PCReq` message, such as:

- `Bandwidth Object`: indicates the bandwidth of the required path.
- `Metric Object`: specifies the metric type to be optimized (e.g., TE metric).
- `Record Route Object` (`RRO`): indicates, in case of re-optimization, the strict route of the established path to be re-optimized.
- `Include Route Object` (`IRO`): indicates specific required nodes/segments to be included within the path.
- `Exclude Route Object` (`XRO`) [10]: indicates specific required nodes/segments to be excluded from path computation.
- `LSP Attribute Object` (`LSPA`): indicates attributes such as desired local protection schemes (e.g., Fast ReRoute).
- `Objective Function Object` (`OF`) [11]: specifies the optimization criteria, namely objective functions, to be applied in the path computation.
- `Synchronized Vector Object` (`SVEC`): allows a PCC to require synchronized path computations mainly for optimization and protection purposes. Three flags are defined to require respectively link, node and Shared Risk Ling Group (SRLG) diverse path computation.

B. PCEP Path Computation Reply message

Upon the elaboration of a `PCReq` message, if no errors occur, the PCE returns a `PCRep` message, containing only one mandatory object, the `RP Object`, which provides the request-ID of the computed request. In case of path computation failure because of unsatisfied set of constraints, the `PCRep` includes a `NO-PATH Object`, possibly augmented with additional information about the failure reasons. For example, if the path could not be established due to insufficient bandwidth, the `Bandwidth Object` is enclosed to notify the PCC. If the PCE successfully computes the required path, the `PCRep` includes an `Explicit Route Object` (`ERO`). The `ERO` may contain a list of strict and/or loose nodes (including IPv4/IPv6 addresses, Autonomous System (AS) numbers as abstract nodes) to be used by the signaling protocol to establish the LSP. In addition, the `Metric Object` may be provided, indicating the metric cost of the computed path.

IV. PCE in Single Domain and Single Layer Networks

Single domain networks are owned and maintained by a single administrative entity, e.g. network provider. Single layer networks are based on a unique switching capability option. The consequence of the aforementioned two assumptions is that the routing protocol can exchange the whole network topology without any restriction or summarization (e.g., within OSPF-TE areas). In this context, if the information flooded by OSPF-TE is used also by the PCE for path computation, the adoption of a PCE could appear as unmotivated. However, there are reasons claiming for the opposite.

Based on the desired QoS level offered to customers, the operator may choose to adopt complex path computation algorithms in order to satisfy multi-constraint Service Level Agreement (SLA) profiles, or a set of synchronized paths that are part of the same service instance (also referred to as path

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PCEP OBJECTS

PCEP object	Message	Description
OPEN	Open	PCEP session parameters
KEEPALIVE	Keepalive	Session liveness
RP	PCReq,PCRep	Request identification and parameters
END-POINT	PCReq	Specifies the LSP source and destination nodes
BANDWIDTH	PCReq,PCRep	Specifies the required LSP bandwidth
METRIC	PCReq,PCRep	Required/Computed metric
LSPA	PCReq,PCRep	LSP attributes (local protection, priority)
NO-PATH	PCRep	Requested path not found
ERO	PCRep	Nodes/link sequence of computed path
SVEC	PCReq	Synchronized computation (diverse link, node, SRLG disjoint)
XRO	PCReq	Exclude nodes from computed path
IRO	PCReq	Include nodes within computed path
OF	PCReq	Path Computation objective function
PCEP-ERROR	Error	Protocol errors
NOTIFICATION	PCNtf	Event notifications
CLOSE	Close	PCEP session termination

computation bundling), or perform a global re-optimization of a set of provisioned paths. The former aspect is discussed in Sec. IV-A, while the latter aspect motivates the adoption of the stateful PCE, discussed in Sec. IV-B.

The PCE can operate on a full topology with a desired level of detail. Additional information may be considered to improve the path computation, depending on the data-plane technology. For example, a number of information parameters typical of the management plane (e.g., signal quality measurements, alarms) can be used by the PCE. Additional methods to provide the PCE TED with updated and enhanced information with respect to standard OSPF-TE information are discussed in Sec. IV-C, concerning the TED update mechanisms.

The path computation of LSPs in specific data-plane networks is the result of a set of complex procedures, performed at different stages and times. This is the case of WSONs, in which path computation includes routing, wavelength assignment, possible physical impairment validation. The adoption of PCE, fully (or partially) centralizing all such functions, may be beneficial and has been deeply investigated in the last years. PCE-based routing and wavelength assignment procedures are discussed in Sec. IV-D, impairment-aware path computation schemes are discussed in Sec. IV-E. Recently, flexible optical networks, also referred to Spectrum Switched Optical Networks (SSON), have improved the degree of freedoms of path computation, including flexibility on the selection of additional transmission parameters, such as signal modulation format and selected spectrum width based on the desired optical reach. The use of the PCE in SSON is discussed in Sec. IV-F.

Besides LSP provisioning, the PCE may be utilized, in the context of GMPLS, to improve network reliability. In Sec. IV-G, PCE-based protection and restoration techniques are reported and discussed.

This section includes the discussion of typical PCE functional aspects that are not only related to a single domain scenario, but that typically occur in it. As an example, path computation bundling or impairment-aware path computation are also possible in multi-domain or multi-layer networks, however, without loss of generality, their utilization is described in the single-domain scenario.

A. Path Computation Bundling

Path computation in large networks requires a careful treatment depending on the kind of services utilized by connections. In particular, services that continuously and dynamically require the provisioning of an LSP may be responsible for resource allocation inefficiency when global optimization is considered. Moreover, for some services (e.g., virtual private networks, VPNs, serving multiple sites), a single instance may require the setup of multiple LSPs (e.g., a full mesh). Nonetheless, high-value services may require protection mechanisms, made available by two (or more) dependent LSPs with specific constraints (e.g., link and node disjoint). To address such requirements, the PCEP protocol allows the bundling of multiple requests onto either a single *PCReq* message.

Two criteria are considered to classify bundled requests. Based on the bundling computation nature, requests can be *synchronized* (in the case many LSPs are required to be computed jointly and simultaneously, optionally subject to a single global optimization parameter, e.g., minimize the maximum link load), or *not-synchronized* (in the case many LSPs can be computed in a serialized fashion). Based on the relationship among bundled requests, they are referred to as *independent* (in the case they are not related each other) or *dependent* (in the case they cannot be computed independently, e.g. diverse path computation). Multiple LSP requests are synchronized within the `PCReq` through the `Synchronized Vector Object` (`SVEC`) [7], [12]. The `SVEC` object also includes optional parameters for dependent path computation (i.e., link/node/SRLG disjointness).

The bundling selection may occur either at the NMS or at PCE, as shown in Fig. 4. In the former case (Fig. 4-a), NMS is responsible for service provisioning and submits synchronized `PCReq`s to the PCE. In the latter case, the PCE receives generic `PCReq`s and is responsible for (optional) service differentiation and bundling operation. In Fig. 4-b a PCE architecture comprising service differentiation and bundling is shown. A number of parameters such as request interarrival time, time spent in the request queue, number of pending requests are considered in the bundling selection algorithm. Typically, bundling selection should achieve a reasonable trade-off between efficient TE performance and setup delay.

The works in [13] and in [14] apply path computation bundling in MPLS networks for distributed Grid Computing applications. The idea is that, as each LSP configuration causes temporary irresponsiveness in commercial routers, bundling together requests is beneficial for both resource allocation and outage time minimization at routers. The requests generated by different source nodes are handled in a queue by an augmented, service-oriented NMS.

In the context of optical networks, the works in [15], [16] propose concurrent path computation exploited by the PCE to perform joint optimization of network and Grid Computing

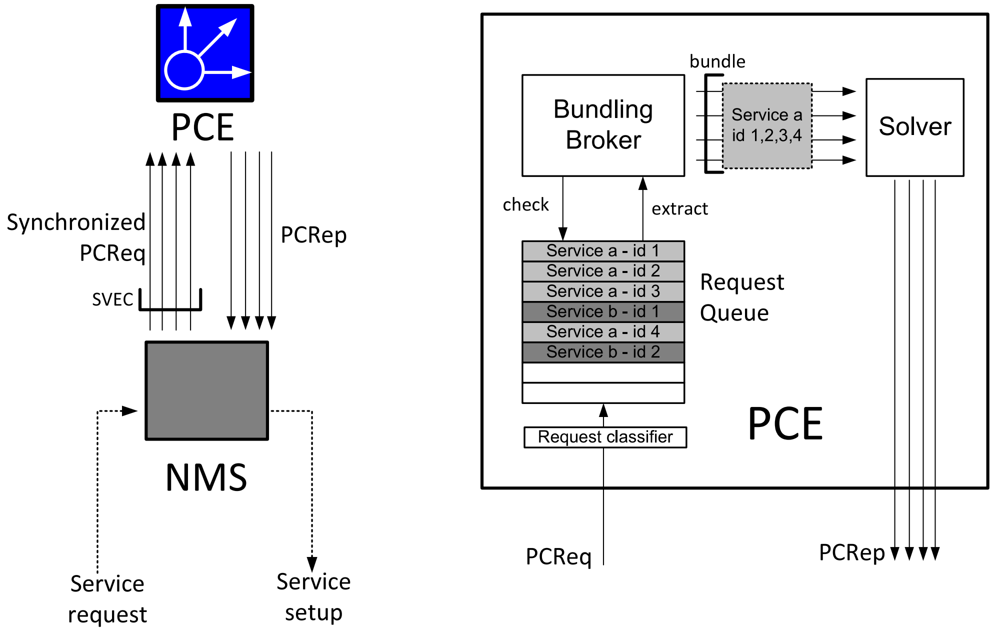


Fig. 4. Path computation bundling: NMS-driven (a), PCE-driven (b).

resources. In particular, PCEP is proposed as the standard interface between the network resource manager and the grid resource manager. Such solution allows joint resource optimization and enables the utilization of legacy control plane implementations, thus avoiding the need for routing extensions just for a specific application.

The work in [17] evaluates bundling effects in WSON scenario with service differentiation. The PCE is equipped with a `PCReq` counter triggering concurrent path computation upon overcoming a given threshold. Results show that blocking probability, defined as the probability that an LSP demand is refused due to unavailable resources, decreases linearly as the counter threshold increases, while the LSP setup time, defined as the time required to perform both path computation and resource reservation, increases exponentially. Recent work in [18] proposes integer linear programming formulation and a comparison with heuristics based on greedy randomized adaptive search procedures.

The work in [19] utilizes a buffering prediction method to trigger the computation bundling having a mild impact on the LSP setup time even in presence of low inter-arrival time requests.

B. Stateful and Stateless PCE

The state of an LSP is defined as the detailed information of the network resources used by the LSP. In particular, the utilized route, the reserved bandwidth and the switching capability information are included in the LSP state. Typically, the state of an established LSP is owned by its source node or by the NMS. The PCE TED just stores per-link resource information and does not manage the association between such resources and the utilizing LSP. When such association is available at the PCE, the PCE is called *stateful*, otherwise is referred to as *stateless*.

Path computation operations at PCE are required either in the provisioning phase and in the re-optimization phase [2]. In the former case the only handling of the TED is sufficient, while the latter needs both the TED and the full state of currently operating LSPs (including associated reserved resources and routes). Re-optimization performed by a stateless PCE requires that each LSP source node (or the NMS) provides the PCE with the state of established LSPs. Such procedure can introduce significant scalability issues during re-optimization. Conversely, the adoption of a stateful PCE makes re-optimization agile and effective.

The stateful PCE has encountered initial concerns due to the higher complexity introduced by the strict synchronization between PCC and PCE, including control plane overhead, race conditions, global path computation complexity. As a consequence, efforts have been placed so far to improve the capability of the stateless PCE to provide effective path computation. In the meanwhile, the need of re-optimization operations has pushed the standardization process towards the co-location of PCC and the NMS. In [20], global concurrent optimization procedures are defined between NMS and a stateless PCE, thus avoiding complex synchronization issues.

In [21], the stateless PCE implementation is augmented with static management information for avoiding resource contentions among subsequent lightpath requests in a transparent WSON. Blocking due to resource contention occurs during the backward signaling phase when two or more lightpaths try to concurrently reserve the same wavelength on the same link [22], [23]. The management TED information is used to protect incoming LSP requests from multiple resource reservations through the estimation of the time required by reservation process. Such solution enables a stateless PCE, simpler to be implemented and managed, to reduce connection blocking due to resource contentions.

In [24], a stateless PCE is enriched with stateful information, called context awareness, in the specific context of disruptive failure events and subsequent restoration. Results show that context awareness limited to one second can provide up to 50% of the overall recovery time.

Recently, a couple of IETF drafts [25] [26], have raised novel interest towards stateful PCE. In fact, PCE-based bundling can be utilized also for global re-optimization. Efficient integer linear programming (ILP) solutions have been demonstrated to be effective in large TE systems, providing both rapid convergence and significant benefits in terms of network utilization [27]. Such solutions require both global visibility of LSP state and the ordered control of path reservations across devices within the controlled system. In [25], a set of PCEP messages extensions have been proposed to manage new functions between the stateful PCE and PCC. These functions include:

- *Capability negotiation*: stateful PCE extensions and capability support are communicated in the `Open` message, upon session establishment.
- *LSP state synchronization*: stateful PCE receives by the PCC the state of an established LSP (i.e., typically, the LSP source node) and stores it.
- *LSP update request*: stateful PCE requires to run re-optimization on selected LSP and performs update requests to related PCC.
- *LSP State Report*: PCC provides the stateful PCE with the updated attributes of an LSP, after LSP modifications.
- *LSP control delegation*: PCC grants to stateful PCE the right to update LSP attributes on one or more LSPs. In this way, the PCE becomes the rights owner of those LSPs. The PCC may then withdraw the delegation. Alternatively, the PCE may release the delegation.

To handle the aforementioned functions, two PCEP messages are defined: the `Path Computation State Report Message` (`PCRpt`) is sent by a PCC to a PCE in order to report the status of one or more established LSPs, and the `Path Computation Update Message` (`PCRpt`) is sent by the PCE to the PCC to update LSP parameters on one or more LSPs upon global re-optimization. A number of novel PCEP objects is identified, among which the `LSP Object`, which univocally identifies the established LSP.

C. PCE Traffic Engineering Database update

The PCE TED stores the updated snapshot of the controlled network. In particular, the TED contains information on the network topology (e.g., node, links and their interconnection) and the current availability of the network links (e.g., available bandwidth, available wavelengths). One debated issue refers to the TED update mechanism. In the distributed approach, each node stores a TED built up by resorting to the OSPF-TE flooding. A node has detailed information (i.e., the full set of resource reservation state) only about local attached links. TE information of the whole network is collected through the exchange of Opaque Link State Advertisement (Opaque LSA). In the PCE architecture different strategies can be adopted in order to keep updated the TED stored at the PCE.

A first solution represents the direct extension of the distributed approach: the PCE directly operates within the distributed control plane and populates the TED resorting to the routing instance of the network. If the PCE is co-located within a node, the OSPF-TE session run by the node is internally extended to the PCE. Conversely, if the PCE is a distinct server, the affiliation is possible by opening a passive OSPF-TE session, in which Opaque LSA are collected but not generated [28]. Authors of [29], [30] adopt this solution in the development of an open-source PCE, evaluating the TED update rate scalability. Moreover, by emulating highly dynamic network scenario, they propose to re-perform path computations directly affected by incoming TED updates, thus reducing the effect of out-of-dated information, at the expense of increased PCE load and path computation delay.

An alternative solution is based on the use of management plane information: the PCE resorts to Simple Network Management Protocol (SNMP) messages directly coming from the network equipment. The NMS is then responsible of both triggering PCC requests and managing TED update. With this solution, the decoupling of the control plane information updates and the TED may facilitate the PCE deployment in networks that implement limited control functionalities (e.g., only signaling) or hybrid networks where node architectures are different and do not support common control plane features [21].

Besides the aforementioned solutions, a number of works propose mixed or intermediate approaches. A combined approach is proposed in the implementation work [31], in which a strict cooperation between control plane and management plane is enforced. In particular the PCE TED entries are updated by both OSPF-TE session peering and SNMP extra monitoring information.

In [32], within a WSON scenario, the TED update function is realized through the exchange of PCEP Notification messages between the source node and the PCE. In particular, two methods are proposed: the *Reactive* PCE and the *Proactive* PCE. In the reactive PCE, the PCEP messages are related to the LSP setup and tear down events triggered by the signalling protocol. In the proactive PCE, the update occurs immediately by the PCE itself after the path computation, assuming that the computed LSP is already established and possible errors occurring during the signalling phase are detected and notified to the PCE by the source node. The proposed mechanism uses detailed information (e.g., wavelength status) that current OSPF-TE implementation do not support, and provides the necessary information set for the implementation of either stateful or stateless PCE. Simulation results show that the proactive PCE significantly reduces blocking due to the latency between path computation and resource reservation, and due to the resource contention during signalling phase (i.e., resources occupied at a node by another LSP between RSVP-TE Path and Resv message).

A similar approach is also considered in work [33], in which the PCE, upon path computation, performs temporary pre-reservation of the computed resources within the TED by means of timers, thus eliminating contentions. The value of the timer is required to be selected properly, in order to not affect the blocking probability due to resources unavailability, especially in highly dynamic scenarios.

D. PCE Routing and Wavelength Assignment in WSON

In WSONs, lightpaths are provisioned by solving the well known *Routing and Wavelength Assignment* (RWA) problem.

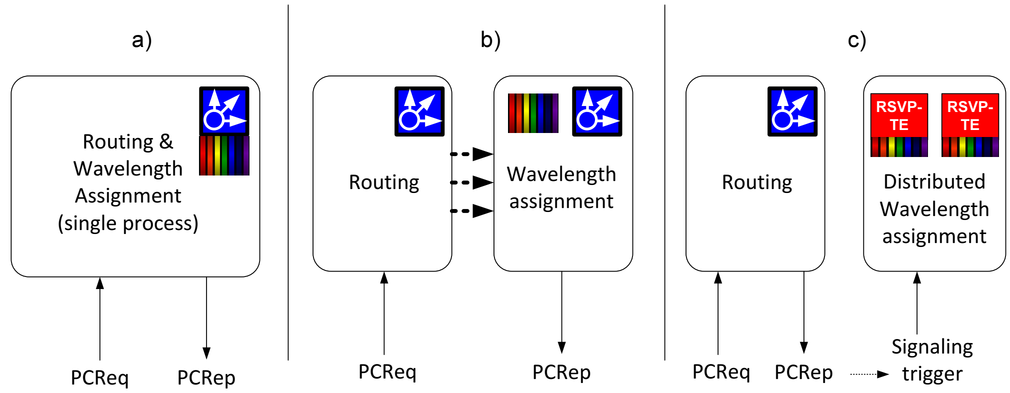


Fig. 5. PCE-based RWA process: R&WA (a), R+WA (b), R+DWA (c).

Optimal joint solutions have been demonstrated to be NP-complete, so a large number of heuristics have been investigated and proposed. In the classical source-based routing scheme, the source node is responsible to perform routing. Such selection is based on topology and TE information that are made available by distributed routing protocols (e.g., OSPF-TE). In the context of transparent WSON, the additional wavelength continuity constraint (WCC) is considered to provision a single wavelength from source to destination. Wavelength assignment is typically performed during signaling through a distributed scheme. The set of available wavelengths is skimmed at each node and carried by the signaling protocol (i.e., by RSVP-TE Path message) up to destination, that selects the wavelength, based on given policies (e.g., first-fit, random-fit). Wavelength reservation is then enforced backwards up to source (i.e., by the RSVP-TE Resv message).

Possible inefficiencies of the aforementioned scheme rely on functional, TE and temporal separation between routing and wavelength assignment. Functional separation derives from the classical GMPLS approach: path computation / routing and wavelength assignment processes are performed independently. This excludes an accurate joint RWA, thus leading to suboptimal solutions. TE separation derives from applying routing and wavelength assignment on different information databases (e.g., R on global OSPF-TE TED, WA on local RSVP-TE path state database). Temporal separation occurs by performing routing and WA at different instants, thus incurring in temporally misaligned databases. In addition, WA is spread in time and intrinsically subject to information desynchronization (e.g., leading to collisions). The introduction of centralized path computation provided by PCE may help to reduce, even remove, such inefficiencies.

In [34], the following PCE-based RWA process alternatives are presented:

- *Combined Routing and Wavelength Assignment (R&WA)*: a single PCE process performs path selection and wavelength assignment (as sketched in Fig. 5-a).
- *Separate Routing + Wavelength Assignment (R+WA)*: two distinct processes (within a single PCE or at two dedicated PCEs) are performed. The first process selects one or more potential paths, then the second process selects the wavelength along with the final selected path. The procedure is sketched in Fig. 5-b.
- *Routing and distributed Wavelength Assignment (R+DWA)*: PCE performs a standard path computation, unaware of detailed wavelength availability. Wavelength assignment is performed along the computed path in a distributed manner through signaling protocols (RSVP-TE). The procedure is sketched in Fig. 5-c.

The three RWA alternatives are illustrated in Fig. 5.

In [35], RWA PCE-based solutions are compared. In particular, R&WA is demonstrated to achieve the best resource allocation performance at the expense of increased computational complexity. A trade-off solution is also proposed performing distributed routing and PCE-based wavelength assignment based on a set of candidate routed paths.

In [36], [37], eight schemes originating from the three general RWA schemes of Fig. 5 are proposed and evaluated. In particular, random and first fit wavelength assignment are compared and a R&WA with priority queue based on path length in term of hops is proposed.

In [38] a PCE-based architecture supporting bidirectional lightpaths is proposed, where the PCE performs R&WA and two different signaling mechanisms are evaluated, employing the upstream label and enhanced label set, respectively.

In [39], an extended R+DWA scheme is described. In particular the PCE provides the route along with a set of candidate wavelengths, including a primary and a set of alternate wavelengths. Alternate wavelengths are utilized in the case of signaling collisions occurring at an intermediate node during the RSVP-TE Resv message flooding.

E. PCE and Physical Impairments in WSONs

Lightpath transparency in WSONs increases physical propagation distances over which physical layer impairments accumulate, potentially resulting in unacceptable lightpath Quality of Transmission (QoT). A number of end-to-end physical parameters are considered which describe QoT, among which signal attenuation, amplified spontaneous emission (ASE), polarization mode dispersion (PMD), chromatic dispersion (CD), self-phase modulation (SPM), cross-phase modulation (XPM).

The intrinsic complexity of impairment-aware routing and wavelength assignment (IA-RWA), due to the large amount of

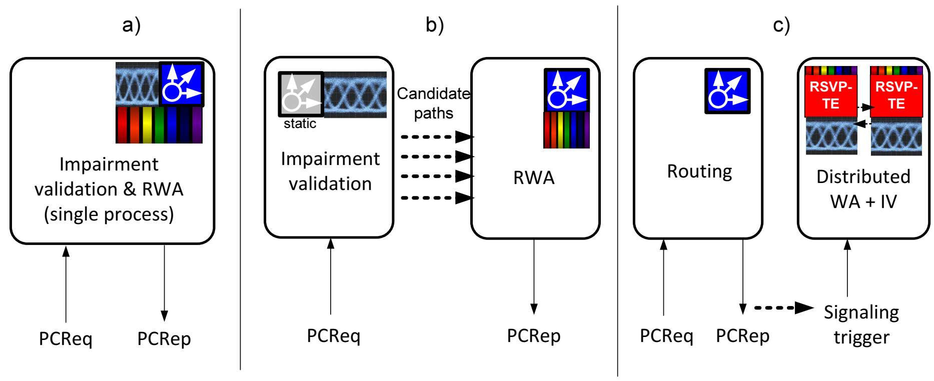


Fig. 6. PCE-based impairment validation and RWA: IV&RWA (a), IV-Candidates+RWA (b), R+DWA+IV (c).

physical information to handle and the computational effort to perform impairment estimation, has driven the proposal of PCE-based architectures for IA-RWA. Thus, PCE may be required to perform RWA while accounting for physical layer impairments in order to guarantee an adequate QoT, e.g. a value of bit error rate (BER) lower than a threshold. Thus, beside RWA, PCE has to validate the impairments and the QoT. This impairment validation (IV) assumes the knowledge of specific physical layer impairment parameters and the adoption of a QoT model [40]–[45], which depends on the WSON scenario, bit-rate, routes, wavelengths, etc.

IETF has proposed several architectures for impairment validation [46], which are hereafter detailed and sketched in Fig. 6:

- *Combined IV and RWA Process (IV&RWA)*: the processes of impairment validation and RWA are aggregated into a single PCE. In this case the same PCE exploits routing and wavelength availability information as well as physical parameters. In this way RWA may be jointly performed accounting for impairments (Fig. 6-a).
- *IV-Candidates + RWA Process (IV-candidate+RWA)*: the impairment validation and RWA processes are separated and performed by two different PCE entities. In this case, the IV PCE provides the RWA PCE with a set of validated *candidate* routes, i.e. each route and each wavelength along these routes guarantee QoT. Thus, the IV PCE, besides routing information exploits physical parameters information. Then, RWA PCE performs RWA on the set of validated candidate routes without accounting for physical parameters and QoT (Fig. 6-b).
- *Routing + Distributed WA and IV*: a PCE, unaware of wavelength availability information and physical parameters, is assumed. Wavelength assignment and impairment validation are performed exploiting either signaling or routing protocol extensions (Fig. 6-c).

This section is focused on impairment-aware PCE architectures, thus IV&RWA and IV-candidate+RWA will be analyzed. With respect to the distributed impairment validation, with IV&RWA and IV-candidate+RWA, only the PCE has to store physical parameters information, while in the distributed case all control plane nodes are required to store physical parameters. Thus, IV&RWA and IV-candidate+RWA relax the amount of information stored in the control plane nodes.

Recent research works studied and implemented QoT-aware PCE architectures. In [47]–[51], IV&RWA PCE architectures are described and demonstrated, especially in the case of WSONs employing 10 Gb/s transmission. In [52], a Double Cooperating PCE architecture is presented and demonstrated. Similarly to IV-candidate+RWA, it is composed of two PCEs, but differently from IV-candidate+RWA, a candidate or a set of candidates are computed by the PCE RWA without accounting for physical parameters. Then, the candidates are passed to the IV PCE for impairment validation which sends back to the RWA PCE the validated candidates. However, this architecture may suffer from increased delay with respect to IV-candidate+RWA if no candidate presents acceptable QoT and the RWA PCE has to compute other candidates.

Works in [53]–[56] are focused on IV&RWA PCE for higher rates or multi bit-rate scenarios (e.g., 40 and 100 Gb/s). In [57], a study and a comparison on several QoT-aware PCE architectures are carried out. In particular, authors show that IV-candidate+RWA architecture may experience a larger delay than IV&RWA because of the communication between IV PCE and RWA PCE. However, based on the algorithm complexity to jointly compute and validate RWA, IV&RWA may experience a larger delay to compute RWA than IV-candidate+RWA. The larger complexity of IV&RWA may be justified in terms of lower achieved lightpath blocking probability. Indeed, with IV-candidate+RWA, the IV PCE has to provide the RWA PCE with candidate routes which satisfy QoT for any wavelength. This is typically done by assuming the worst-case scenario (e.g., maximum cross-phase modulation). However, the worst-case approach of the IV-candidate+RWA may estimate some source-destination pairs as unreachable, thus causing lightpath blocking even if the worst-case scenario does not occur. Conversely, IV&RWA performs a more accurate impairment validation (e.g., estimating the actual cross-phase modulation) and overcomes these problems of blocking. Generally, if wavelength-dependent impairments (e.g., cross-phase modulation) are particularly relevant in a network scenario (such as a multi-rate scenarios) IV&RWA may present better performance than IV-candidate+RWA in terms of blocking probability, but it requires a larger delay for path computation. IV-candidate+RWA may achieve even lower path computation delays if the set of candidate routes are pre-computed by the IV PCE and stored by the RWA PCE. In this way, the communication between the two PCEs occurs off-line and, upon computation request, the RWA PCE may promptly exploit candidate route information.

Work [58] enhanced the IV-candidate+RWA architecture by proposing an extension, called `Guard Band`, to be associated to each route in the set of candidate routes. The extension permits to obtain similar blocking probability of IV&RWA in a multi-rate scenario degraded by cross-phase modulation because the extension permits to avoid the consideration of the worst-case scenario.

F. PCE in Flexible Spectrum-Switched Optical Networks

Recent works have also studied PCE in emerging flexible (or elastic) optical networks, which are enabled by the availability of bandwidth variable wavelength selective switches (i.e., capable of switching configurable portion of the spectrum) and flexible transponders (i.e., capable of dynamically selecting bit-rate and modulation format). In such networks [45], [59], [60], instead of a single wavelength, lightpaths occupy a variable portion of the spectrum, called *frequency slot* [61], whose width depends on the bit-rate and the selected modulation format. Consequently, in flex-grid optical networks (also referred to as *spectrum switched optical networks*—SSON), the problem of RWA is replaced with the problem of routing and spectrum assignment (RSA). Besides path computation, RSA consists in assigning a frequency slot, which is defined by a *central frequency* (i.e., the carrier) and by a *slot width* (i.e., the required bandwidth). Therefore, PCE has to be designed to perform spectrum assignment, i.e. to assign a central frequency and a slot width, instead of an only ”wavelength”. Also PCEP should be extended for flex-grid in order to carry information related to the assigned frequency slot (e.g., in the `PCRep` message). A frequency slot format has been proposed within IETF [61]. Moreover note that, similarly to the RWA problem, routing and spectrum (or frequency slot) assignment may be jointly computed (*R&SA*), may be separated (*R+SA*), or may be separated with distributed spectrum assignment (*R+DSA*) [61].

Another important consideration lies on the fact that, differently to WSONs where the modulation format is a constraint of the source node (the PCC), in flex-grid optical networks, the modulation format may be an output of the RSA [62], [63]. Indeed, RSA is strictly related to the impairment validation and the modulation format selection. In particular, given the bit-rate, the more efficient the modulation format in terms of occupied spectrum, the less robust in terms of QoT. As an example, 16-quadrature amplitude modulation (16-QAM) halves the required bandwidth with respect to quadrature phase shift keying (QPSK), but 16-QAM presents a limited optical reach with respect to the other one. Thus, because of the QoT, modulation format selection is affected or affects path computation. Then, frequency slot assignment, in particular the slot width, is affected by the modulation format selection. Thus, in [63], PCEP is extended to include modulation format information, not only in the PCEP `PCReq` as in classical WSONs, but also in the `PCRep` message, given that modulation format selection is an output of RSA. Works [63] and [64] report experimental results on the aforementioned proposed approach evaluated in optical network testbeds employing coherent detection applied on single carrier and orthogonal frequency division multiplexing (OFDM) signals, respectively.

Finally, a thematic in flex-grid optical networks is the de-fragmentation problem. Especially in dynamic networks, after some lightpath release, the spectral efficiency is compromised due to the fragmentation of the available spectrum into small noncontiguous spectral bands, decreasing the probability of finding sufficient contiguous spectrum along the whole route for new lightpath requests. However, thanks lightpath re-routing or spectrum re-assignment, it is possible to reduce the fragmentation (*defragmentation*) and to improve the spectral efficiency. Thus, several defragmentation (i.e., re-optimization) solutions have been recently proposed for flex-grid optical networks [65]–[68] and may be applied to a PCE scenario. In [69], PCEP is extended to support defragmentation by introducing two new messages called `Spectrum Defragmentation Request Message` and `Spectrum Defragmentation Reply Message`.

G. PCE-based protection and restoration

The GMPLS control plane is also responsible for network reliability. Indeed, since each link in a WSON can be traversed by tens of wavelength channels, a single failure (e.g., link failure) can generate a huge loss of data. Two different types of reliability mechanisms have been mainly proposed in literature for WSONs: protection and restoration mechanisms. In the former type, backup bandwidth is reserved upon connection establishment. In the latter type backup bandwidth is reserved only upon failure occurrence.

Most research work on protection mechanisms aims at reducing the amount of reserved backup bandwidth. In particular *shared path protection*, introducing backup bandwidth sharing among lightpaths that cannot be disrupted by the same failure, emerged as the most promising protection mechanism. Several distributed implementations of shared path protection have been proposed [70]–[72]. However, a centralized PCE storing the working and backup path of all established lightpaths, can help in achieving higher sharing ratio and improve resource utilization.

In [73] a shared path protection scheme is proposed using the PCE. Thanks to the centralized path computation an improved sharing ratio is achieved finally resulting in a reduced lightpath blocking probability.

In [74] a PCE implementation performs shared path protection in WSONs by applying a modified 2-step Dijkstra algorithm and extending PCEP accounting for protection type, same working/backup wavelength indication and assignment method selection.

The work in [75] proposes bulk path computation encompassing shared path protection in WSONs. Wavelength sharing strategies are demonstrated to noticeably benefit from dependent PCE-based bulk computation by means of a flexible version of the greedy adaptive search procedure (GRASP).

Research work on restoration mechanisms mainly considered two performance parameters: recovery blocking probability and recovery time. Indeed, if restoration is used, a burst of messages is typically generated on the control plane upon failure occurrence because all the disrupted lightpaths try to be restored as fast as possible. Recovery blocking probability is therefore degraded because a number of recovery attempts may be blocked due to resource contention [76]. Therefore a centralized PCE can be effectively used for coordinating the recovery attempts achieving to significant reduction of blocking due to resource contention [32]. On the other hand, the PCE utilization implies additional PCEP communications and can significantly degrade the recovery time. This trade-off has to be carefully considered when proposing the utilization of PCE for restoration.

In [77] a recovery module is added to the PCE, that acts as manager of the whole set of recovery processes (i.e., fault notification, segment path restoration, signaling triggering) in order to reduce the overall recovery time. A hierarchical architecture is considered, also suitable for inter-domain LSP recovery. Such architecture allows the aggregation of fault notifications, thus reducing signaling storms events. Recovery times below 50ms are achieved.

In [78] two PCE-based dynamic lightpath restoration methods are proposed. The first method relies on heuristics aiming at reducing resource contention. In particular, after the failure notification, each restoration request is computed sequentially and selected links are overweighed in order to discourage utilization of future restoration requests. The second method employs bulk ILP-based computation. PCE-based rerouting experiences beneficial effect in terms of connection blocking probability at the expenses of setup time, while wavelength suggestion is preferable to be enforced through signaling.

In [79] a PCE-based restoration is implemented on a transparent WSON. The PCE receives restoration requests enclosing Exclude Routing Object (XRO) of the entire failed path or failed elements in the case such information is available at the PCC. Centralized wavelength assignment is also evaluated, however performance results show that distributed assignment is more effective. This is due to the ability of distributed schemes to better collect updated wavelength availability information. Restoration show a time range up to 140 ms.

Similar achievements are reported for translucent 4-node WSON testbed in [80], in which the PCE-based restoration accounts also for accumulated optical signal-to-noise ratio (OSNR) pre-validation and 3R regenerators node service. In such a context disruption time ranges are in the order of 300 ms.

V. PCE IN MULTI-LAYER NETWORKS

In the context of GMPLS, a multi-layer network (MLN) consists of transport nodes with interfaces operating at multiple data plane layers of either the same or different switching technology and controlled by a single GMPLS control plane instance [81]. Each node interface switching technology is identified by a specific Interface Switching Capability (ISC). Examples of ISC are Packet Switching Capable (PSC) or Lambda Switching Capable (LSC). In MLNs, a Label Switched Path (LSP) starts and ends in the same layer (i.e., ISC), and may cross one or more lower layers. Once an LSP is established within a layer from one layer border node to another, it can be used as a data link in an upper layer. Furthermore, it can be advertised as a Traffic Engineering (TE) Link and exploited in the path computation of LSPs originated by different nodes. Such TE Link is referred to as Forwarding Adjacency LSP (FA-LSP). A FA-LSP has the special characteristic that it does not require the set up of a routing adjacency (peering) between its end points. A Virtual TE Link (VL) is defined as a lower-layer LSP that is advertised to the upper layer as it were fully established (i.e., as an FA-LSP), but it is not established [81] (i.e., it is just computed but it is neither signalled nor cross-connections are performed). Fig. 7 shows a MLN with an example of a FA-LSP (and the related installed lower-layer LSP) and a VL.

As specified in [82], the support of VL does not require any GMPLS routing extension. Thus, both VLs and FA-LSPs are advertised to the upper layer as TE links without distinguishing them. If an upper-layer LSP is set up by utilizing a VL, the underlying LSP must be immediately signaled in the lower layer. Signaling can start either dynamically (i.e., triggered by the signaling at the upper layer) or upon specific configuration performed by a management entity (e.g., Virtual Network Topology Manager – VNTM). The latter solution is the preferred one by network operators since it allows the control of the lower layer, thus preventing an unconditional and frequent LSP set up or tear down (i.e., network instability). Some of the motivations that drive the possible set up of VLs in place of pre-established LSPs, are the following. First the pre-provision of lower layer LSPs may be disadvantageous since it might reserve bandwidth that could be used for other LSPs in the absence of upper-layer traffic. In addition the utilization of VLs makes the lower layer data and control plane more stable, since it avoids to pre-provision LSPs that in the absence of upper-layer traffic could be torn down for re-optimization purposes. With VLs, the re-optimization implies just the path computation and it does not trigger any LSP set up or data plane modification. This is particularly important for example in the LSC data plane where node configuration (i.e., OXC cross-connections) may affect the optical quality of transmission of active LSPs, e.g., inducing cross-talk on adjacent channels or triggering complex optical power equalizations.

At the upper layer, the combination of the FA-LSPs and the VLs defines the Virtual Network Topology (VNT) provided by the lower layer [81]–[83]. The VNT facilitates the path computation of LSPs in MLN since it describes the resources at a single layer (both actually available through FA-LSPs and potentially available through VLs) [84]. In this way, a PCE per layer can rapidly perform LSP path computations without considering the large TED including the detailed resources available in the whole MLN, necessary in a single multi-layer PCE. The VNTM is defined as the functional element that manages and controls the VNT [83], e.g. it configures the set up and tear down of VLs. Thus, the cooperation between the VNTM and the PCEs allows the effective set up of LSPs in MLN. Such behavior has been clearly discussed and demonstrated in [85], [86]. Fig. 7 shows a PCE per layer

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

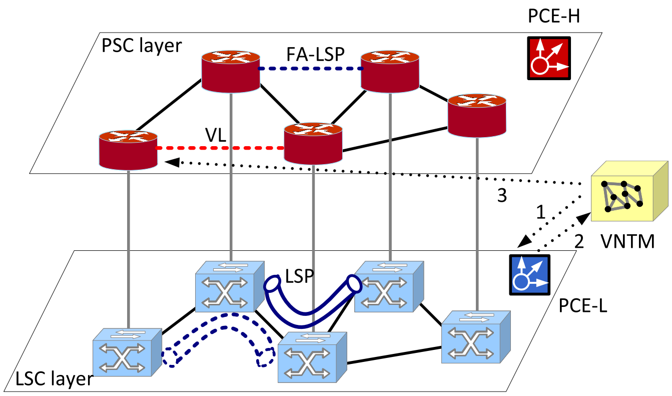


Fig. 7. Double PCE architecture with VNTM in multi-layer networks.

and the VNTM. In particular, the set up of a VL is performed by VNTM through a path computation request to the the LSC-layer PCE (see dotted lines 1,2,3).

The VNTM- and PCE-based architecture enable also the practical utilization of the grooming and multi-layer provisioning strategies largely investigated in the last 15 years (see for example [87], [88]). In [89], [90] such grooming strategies are re-discussed and applied in a PCE-based scenario, i.e., accounting for both efficiency in network resource utilization and implementation aspects (computation time, dynamicity of PCE TED info, amount of PCEP requests). In [91] single- and multi-PCE schemes with/without VNTM are compared and evaluated in terms of computation time, showing that inter-layer computation is achieved in a few tens of milliseconds. In [92], VNTM is extended to enable adaptive advance reservation in carrier-grade Ethernet over WDM networks, as well as to control and manage transit multi-domain paths, by triggering modification of tunnel capacity.

VI. PCE in Multi-Domain networks

TE-based path computation solutions are recognized to provide effective intra-domain network resource utilization. In (G)MPLS networks [1], the routing protocols (e.g., OSPF-TE [93]) are responsible for TE information flooding. In single domain networks, the dissemination of such information within the network is fully allowed. However, in multi-domain networks, this is inhibited by the difficulty of preserving control plane scalability and confidentiality (in multi-carrier scenario) across domains, thus limiting the amount of information exchanged among domains.

Mainly due to scalability problems, multi-domain routing protocols (e.g., Border Gateway Protocol, BGP) only exchange reachability information without detailing the network resource availability, e.g., wavelength availability in WSONs. Hence, effective TE-based path computation strategies are inhibited for inter-domain LSPs, thus strongly impacting the overall network resource utilization [94], [95]. The PCE architecture is able to extend the TE-based path computation in scenarios where resource information flooding is limited (e.g., multi-area, multi-layer, multi-domain) [96].

In multi-domain networks, a single PCE is responsible for path computation inside each domain, while the inter-domain path computation is achieved by means of a coordinated communication process among PCEs, using the procedures defined by PCEP. As a consequence, the knowledge of an adequate amount of information about the network resources of all the involved operators' domains is a key issue. On the one side, an operator should disclose enough information about his network to enable other operators to use its resources. On the other side, an operator usually wants to keep the real network topology hidden in order not to reveal business critical information. For the same reason, TE metrics and parameters usually disseminated in single domain scenarios (e.g., hop count and available bandwidth) are not suitable.

The work [97] discusses cooperative approaches to inter-domain path computation based on PCE, distinguishing between model-based approaches and ad-hoc approaches. Model-based approaches assume the domain sequence is a-priori known and are essentially based on a multi-stage decision problem. The reference model-based approaches are the per-domain path computation (see Sec. VI-A) and the inter-PCE backward path computation, standard (see Sec. VI-B1) and recursive (see Sec. VI-B2). Domain sequence selection represents a sub-problem that is discussed in Sec. VI-C and its most investigated solution, employing the Hierarchical PCE architecture, is detailed in Sec. VI-D.

A. Per-domain path computation

The Per-Domain (PD) path computation technique is proposed in [2] as evolution of the distributed inter autonomous system LSP setup procedure explained in [98]. PD is based on multiple path computations performed during the signalling phase of inter-domain LSP setup. Each domain is assumed to have a PCE responsible for path computation exclusively inside that domain. The sequence of PCE/domains is assumed to be known in advance. In particular, each PCE, given a destination domain, is aware of the next domain hop. In Fig. 8 the PD procedure is sketched. The source node A asks local PCE1 for LSP path computation having destination node Z. PCE1 returns an ERO comprising a subset of strict elements inside the local domain (e.g., A,G,H,P nodes) and a set of loose elements (e.g., destination node Z). The signalling instance (e.g., performed through the RSVP-TE protocol) proceeds until the boundary nodes (BN) of the next domain is reached (e.g., node P). Then, the boundary node asks the ERO expansion to the next domain PCE, which returns the strict ERO (i.e., Q,R,S,O,W nodes), towards the next boundary node. Thus, the signalling phase and the path computation phase are interlaced and the overall inter-domain path is computed and signalled in a distributed way through independent and uncoordinated partial path computations performed by each PCE of the involved domains. The PD procedure, due to uncoordination between domain PCEs, leads to sub-optimal path computation (e.g., the path computed in Fig. 8 has a hop-count metric of 10, while the shortest path has metric 9) and is subject to possible signaling error events. In the latter case, crankback routing (i.e., additional path computation request followed by fresh signaling) may be performed by a boundary node, however it may significantly delay the overall LSP setup time.

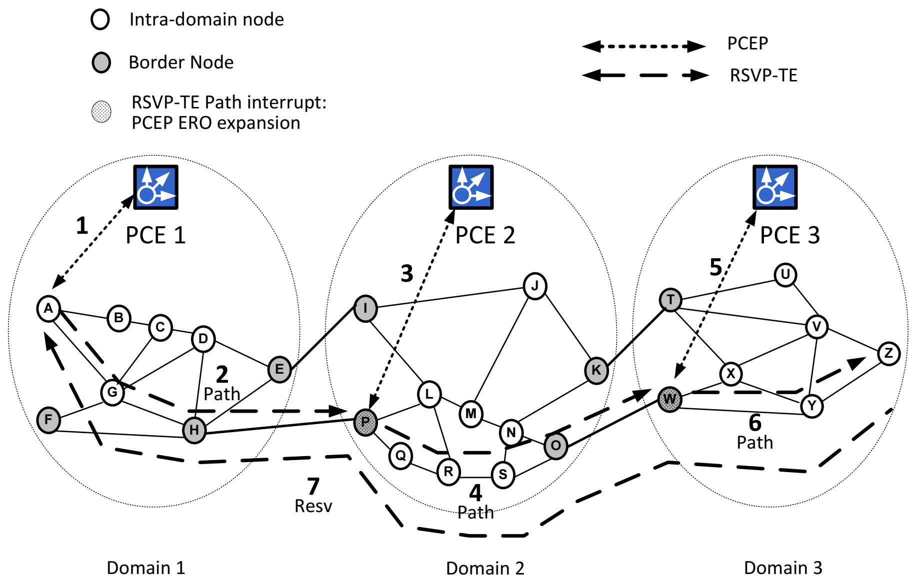


Fig. 8. Per-domain path computation procedure.

B. Inter-PCE path computation

In the Inter-PCE path computation, coordination among PCEs of different domains is introduced. This strategy enables the decoupling between path computation and signalling. Each PCE computation is performed based on the segments computation results provided by the contiguous PCE through PCEP. Among the possible strategies enabling inter-PCE path computation, the *backward* computation is the most considered one, in two versions, standard and recursive. Both versions assume that, for a given source-destination couple, the PCE/domain sequence is a-priori determined. In particular, each PCE is aware of the next involved PCE.

1) Standard Backward Path Computation: The standard backward path computation [96] is exploited through a chain of PCEP sessions between adjacent PCEs. The source and the destination PCEs are those referred to the domain containing the LSP source node and destination node, respectively. Fig. 9 shows the standard backward procedure.

The sequence of involved PCEs (i.e., domains) is a-priori determined. The source PCE (i.e., PCE1), upon path request coming from a controlled node (e.g., node A requesting a path to node Z) triggers a `PCReq` towards the next PCE (i.e., PCE2) and, in turn, the `PCReq` is forwarded until the destination PCE (i.e., PCE3) is reached. The destination PCE computes a path between the destination node and one of the BNs connecting its domain to the domain of the requesting PCE (e.g., segment W-Y-Z). The path segment ERO is enclosed within a `PCRep` and provided to the penultimate PCEs (i.e., PCE2), that computes a path between one BN of the requesting PCE and the BN indicated by the provided segment (i.e., segment I-L-M-N-O-W). The ERO segment is attached to the previous one and the operation is repeated until the source domain is reached. The overall path computation is the result of independent segment path computations because the choice of each BN is delegated to only one PCE. As a consequence the overall computation is sub-optimal (e.g., the path computed in Fig. 9 has hop-count metric equal to 11, in contrast with the shortest 9-hop path), however the coordination between PCEs reduces possible crankback events during the subsequent signalling phase. The work [99] utilizes such technique and proposes three schemes to select the most suitable upstream PCE, based on round-robin scheduling, least-response delay selection and path computation latencies, respectively.

2) Backward Recursive PCE-based Computation (BRPC): BRPC procedure [100] is an inter-PCE path computation technique that computes a constrained shortest path, in a reverse fashion, from the destination domain towards the source domain. According to the domain path, it is possible to define the entry (exit) BNs as the set of BN that are connected to the upstream (downstream) domain.

To compute the inter-domain path, it is assumed that a consistent metric is used in each domain. The BRPC mechanisms is illustrated in Fig. 10. The PCC (i.e., node A) requests an inter-domain path computation to the source PCE (i.e., PCE1), by sending a `PCReq` message. The `PCReq` is forwarded to the PCEs along the domain path through the client-server chain with the BRPC flag set (in the `PCReq` RP object). The destination PCE (i.e., PCE3) computes a tree of potential paths, referred to as *virtual shortest path tree* (VSPT), from the entry BNs in its domain to the destination node. VSPT paths are included in a strict or loose fashion (i.e., by indicating only the entry BNs) in the ERO object (i.e., branches T-V-Z and W-Y-Z). The corresponding segment weights are included in the `Metric` object of `PCRep` message. The `PCRep` message is then forwarded to the upstream PCE, as shown in Fig. 10. In turn, each PCE computes the VSPT from its own entry BNs to the destination by stitching downstream PCE VSPT branches and forwards the `PCRep` message with updated information to the upstream PCE. As an example, PCE2 computes I-J-K-T-V-Z and P-Q-R-S-O-W-Y-Z branches (note that it is not

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

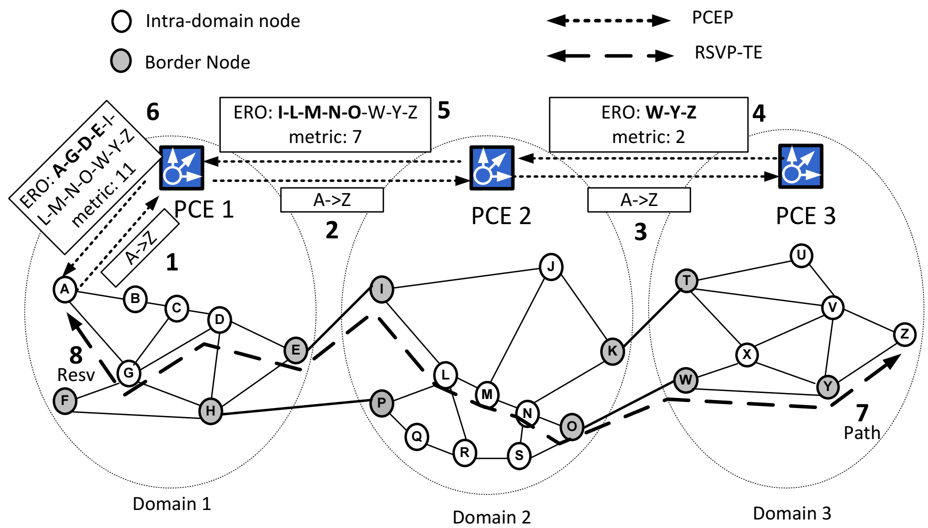


Fig. 9. Inter-PCE standard backward path computation procedure.

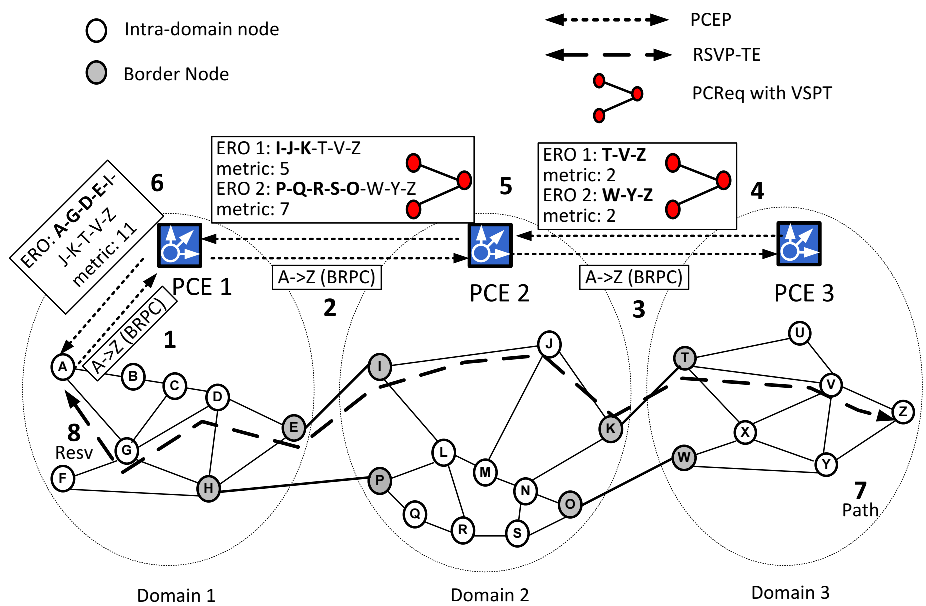


Fig. 10. Inter-PCE Backward Recursive PCE-based computation (BRPC) procedure.

mandatory to stitch both branches provided by PCE3). On receipt of the PCRep message, the source PCE computes the end-to-end inter-domain path using the VSPT and returns the result to the PCC. From the example of Fig. 10, the A-Z path with the shortest accumulated metric is computed (i.e. hop-count metric 9), with respect to the analog example of standard backward computation of Fig. 9 (metric 11) and PD of Fig. 8 (metric 10). This is due to the VSPT flooding mechanism, which allows the selection of BNs by the upstream PCE based on alternative cumulated metrics. The BRPC application achieves optimal path computation on the considered metric.

BRPC procedure avoids the sharing of domain-related information: only aggregate information about the possibility and the weight to reach the destination from a BN is flooded. Also, note that BRPC procedure does not indicate the algorithm to use for the path computation.

The beneficial effect of the BRPC application in large multi-domain topologies with respect to distributed per-domain approach is demonstrated in [101] in terms of resource utilization. In addition, time setup is also reduced since crankback events typically occurring in the per-domain strategy are avoided.

First experimental demonstration in [102] implements BRPC in a multi-domain testbed with control plane and automatic configuration provided by the management plane.

In [103] and [104] standard backward, BRPC and hybrid schemes including domain selection are evaluated, confirming that BRPC represents a good candidate for end-to-end LSP provisioning.

Work [105] utilizes BRPC in multi-area MPLS networks and proposes a load balancing PCE selection technique to improve throughput and path computation latency.

Modifications to BRPC are proposed in [106] to enable inter-domain multipath routing. In particular, for each ingress and egress BN pair, k link-disjoint paths are computed and enclosed within the VSPT, so that they can be suitably (i.e., totally or partially) selected by the source domain PCE.

Authors in [107] propose domain sequence selection based on combined load balancing metrics and BRPC.

In [108] and [109] BRPC is experimentally evaluated for impairment-aware path computation in transparent and translucent WSONs.

Authors in [110] propose to extend PCEP in order to address the wavelength continuity constraint in WSON scenario. In particular, they propose to provide `PCReq` with *LabelSet*, the *Suggested Label* and, for bidirectional LSPs, the *Upstream Label*, as defined in the OSPF-TE and RSVP-TE protocols specifications. The enhanced BRPC performs path computation by computing and updating VSPT taking account the wavelength availability of each sub-tree and eventually pruning those that do not guarantee wavelength continuity along the path.

C. PCE-based Domain Sequence Computation

The inter-PCE procedures described in Sec. VI-A, VI-B1 and VI-B2 assume that the domains sequence to be traversed is known in advance. Since, in complex multi-domain networks, the selection of the domain sequence may strongly affect the overall network performance, research and standardization bodies are currently working on defining the procedures combining domain sequence computation and path computation.

Five main methods have been considered to provide the PCE with multi-domain information for domain sequence computation. Four methods are described in this subsection, while the fifth (Hierarchical PCE) is detailed in the next subsection.

The first method resorts to the information typically announced through the main instance of BGP. BGP does not advertise TE bandwidth information and multiple alternative routes, thus providing the same route indications (i.e., deterministic sequences of domains) which may rapidly induce link congestion. Moreover, BGP is affected by well-known routes flap and oscillations.

The second method, proposed in [111], introduces a service plane dedicated to the exchange of inter-domain TE information. Service plane strictly interacts with the management and the control plane. The notion of inter-AS GMPLS-TE service as a composition of service elements is introduced. Inter-AS provisioning is automated focusing in particular on service composition and activation blocks. This method has been applied in [112] to enable the provisioning of inter-domain point-to-multipoint LSPs.

The third method resorts to the information set advertised through the External Network-to-Network Interface (E-NNI) defined by the Optical Interworking Forum (OIF) [113]. OIF E-NNI Routing has been introduced to address the multi-domain single-carrier scenario. It provides a hierarchical implementation of the OSPF-based routing within the ASON architecture. Similarly to single-domain OSPF-TE procedures, domains and inter-domain links are advertised by routing controllers, one for each domain. Such information, expressed as TE-LSA messages, is flooded to all domain controllers. Such method is employed and extended in [114], [115]. Domain summarization techniques for TE flooding are also investigated in [116], [117].

The fourth method, proposed in [118], [119], [120] and [121], introduces direct or soft interaction between PCEP and path-vector routing. The first solution, employed also in [107], proposes PCEP extensions accounting for path-state routing information. The second and the third solution propose, respectively, peer and hierarchical BGP-like instance devoted to inter-domain links resource advertisement. The hierarchical solution is extended in [122] by means of a dedicated protocol called Domain Sequence Protocol. The fourth solution proposes a path-vector routing at the PCE. One of the advantages of these architectures is that reachability and TE parameters are expressed in terms of aggregated path-based information, thus preserving scalability and confidentiality, and hence suitable for inter-carrier provisioning scenarios. On the other hand, scalability of control plane load and routes oscillations is an issue, so that these solutions are feasible for multi-domain networks with restricted number of domains.

The fifth method is currently the most investigated and employs the Hierarchical PCE architecture, described in the next subsection.

D. Hierarchical PCE

The Hierarchical PCE (H-PCE) [123] architecture proposal defines the procedures for combining end-to-end path computation with an effective domain sequence computation. In the H-PCE, a single parent PCE (pPCE) is responsible for inter-domain path computation, while in each domain a local child PCE (cPCE) performs intra-domain path computation. The pPCE resorts to the Hierarchical TED (H-TED) that stores the list of the domains and inter-domain connectivity information (e.g., inter-domain links with wavelength availability information), to determine the sequence of domains. Moreover, in order to perform more effective inter-domain path computation the pPCE is allowed to ask cPCEs for the path computation of the several edge-to-edge segments of inter-domain LSPs.

In Fig. 11, a path computation request from source node A is forwarded to cPCE of domain 1. Since the destination node does not belong to domain 1, cPCE1 forwards the request to the pPCE. Then, pPCE resorts to the summarized topology in its H-TED and selects the involved domains to be crossed. To retrieve the domain segments, pPCE performs parallel path computation requests to involved cPCEs. Such requests can be single or multiple request, depending on the level of intra-domain summarization provided by H-TED. In the case of

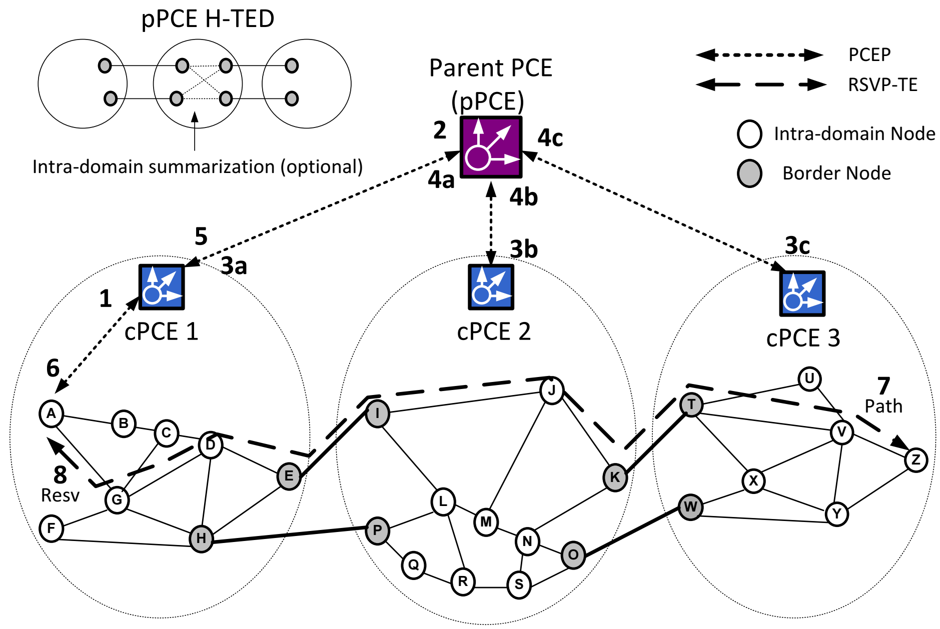


Fig. 11. Hierarchical PCE architecture.

no summarization, pPCE asks for each segment combination. In the example of Fig 11, pPCE asks cPCE2 for I-K, I-O, P-K, P-O segment computation. In the case intra-domain topology summarization is applied, each domain is described through weighted abstract topology (e.g., one node, full-mesh of border nodes, star topology). In the example, if full-mesh summarization is utilized and shortest hop count is considered as weight metric, the virtual intra-domain link I-K has weight 3, shortest among the others. Hence, pPCE will ask only the I-K detailed computation. Segment computations are returned to pPCE, which combines obtained results and compute the end-to-end path, returning the result to cPCE1.

Authors of [124]–[126] have recently developed and tested the multi-domain H-PCE in a WSON lab-trial. The pPCE TED is populated through passive elaboration of OSPF-TE Link State Advertisement (LSA) updates. A number of PCEP extensions is proposed: a PCE ID, a domain ID mapping and the reachability address prefixes range of the domain controlled by the PCE (i.e., to allow endpoint localization). The benefits of the parallelization of the Parent-Child requests are quantitatively evaluated in [124]. While the effectiveness of aggregating the domain topologies as dynamic virtual meshes are evaluated in [125]. In both cases, results on a real topology of 14 optical cross-connects (OXC) and 4 domains show that the path computation delivery is completed in the 3-6 ms time range. Finally, a distributed multi-platform control plane testbed has been developed in [126], validating the aforementioned inter-operability extensions and evaluating two different pPCE domain computation algorithms.

Authors of [127] early proposed the utilization of a hierarchically distributed path computation server (HDPCS) that reduces the setup delay and conducts flexible routing applicable in both multi-layer and multi-domain GMPLS-based networks. The paper compares the performance of the HDPCS with OSPF-based routing and demonstrate the applicability of a centralized hierarchical approach.

However, some recent works sustained possible scalability issues of H-PCE due to PCEP traffic that could overload the control plane, these work proposed therefore several trade-off strategies between resource utilization and generated PCEP traffic [128]–[130].

Authors of [128] propose an algorithm to reduce the number of evaluated path by considering only k random paths in parent PCE. The simulation indicates that the proposed algorithm outperforms previous methods in term of blocking probability and resource utilization. As k increases, the proposed algorithm achieves lower blocking probability and provides improved resource utilization.

Authors of [129] propose a hybrid path computation procedure based on the H-PCE architecture and BRPC. Extensive simulation results show that the proposed approach performs better than H-PCE in terms of network control overhead.

In [130], simulations show that a good trade-off between blocking and control plane load can be achieved using a parent PCE providing only the sequence edge nodes. However, in the more recent work [131], the same authors demonstrate that also a parent PCE providing the detailed list of nodes is a scalable solution applicable in large networks. Moreover this paper shows that, in wavelength continuous WSONs, parent PCE should be enabled to consider wavelength continuity along the whole path. Conversely, providing only the sequence edge nodes, as proposed in [130], achieves the best performance giving to the child PCEs the freedom of choosing the best intra-domain path satisfying wavelength continuity. Both the aforementioned works also estimate the latency between path computation and actual resource reservation, demonstrating that acceptable values are guaranteed by all the proposed schemes.

The work [33] performs a comparison of per-domain, BRPC and H-PCE in terms of blocking probability and a validation of temporary reservation mechanisms for reducing resource contentions similar to the one proposed in [32]. The paper concludes that H-PCE has better performance in terms of blocking probability and its scalability is better than the BRPC mechanism.

The utilization of H-PCE has been proposed also in the MPLS context to enable adaptive advance reservation of inter-domain bandwidth-guaranteed tunnels [132]. In particular, two novel functions are added to the H-PCE framework, namely the *transit tunnel manager* and the *transit topology optimizer*. The former function, deployed at each cPCE, maintains the inter-domain tunnel database and is responsible of notifying updates to the pPCE and triggering tunnel bandwidth change reservation. The latter function, deployed at the pPCE, is responsible of optimizing inter-domain performances by collecting statistics from pPCE TED and perform TE-based load balancing among tunnels. Simulations results show the possibility of maintaining stable transit paths while adapting their reserved bandwidth, thus reducing blocking probability.

E. Multi-domain PCE-based Protection and Restoration

Several works in literature investigated the support of protection and restoration mechanisms in multi-domain scenario. In particular, distributed mechanisms have been proposed for both protection [133], [134] and restoration mechanisms [135], [136]. However, also in this case, the utilization of a centralized PCE in each domain and the proper PCEP communication strategies described in the previous sections may offer significant improvement, especially for protection schemes. This improvement is specifically demonstrated by the work in [134] that performs a comparison between distributed and centralized implementation of protection in multi-domain WSONs.

In [137] 2-disjoint path computation exploiting BRPC is discussed along with proposed extensions to improve PCEP VSPT description and size.

The work in [138] considers the Path Computation Flooding (PCF) approach as a possible BRPC extension for computing optimal end-to-end inter-domain paths without requiring a pre-configured domain sequence. Since PCF presents major scalability issues in terms of network control overhead and path computation complexity, [138] also proposes two novel mechanisms which drastically reduce the control overhead while keeping the connection blocking probability close to the optimal values.

Similarly, [139] proposes an enhanced PCE-based parallel approach (EPA), based on a not differentiated ingress border nodes pair strategy to decrease the computation complex and communication overhead. Results show that the EPA scheme effectively reduces the computation and communication overhead.

The work [140] discusses and compares survivable inter-domain path computation methods. In addition authors propose a method that computes the working path in the forward fashion and backup path in the backward fashion.

The work [141] proposes a multi-domain fast-reroute restoration scheme that uses the PCE for reducing resource contentions due to multiple concurrent LSPs rerouting.

In [142], in-line with PCE-based restoration mechanisms discussed in Sec. IV-G, a 4-domain PCE-based restoration implementation with BRPC is evaluated. Restoration is performed by requesting the restored path by excluding (through XRO object) resources affected by failure. Computation accounts for optical signal-to-noise ratio (OSNR) accumulation and 3R regenerators availability. Path computation latency introduced by only PCE is up to 60 ms and service disruption times are in the order of one second. The authors observed that, due to stateless PCE resorting to simple OSPF-TE TED, in the case of multiple failures, concurrent restored LSPs may collide while trying to reserve the same resources.

In [143] a stateful PCE able to perform multi-fault localization is proposed. Fault notification alarms are sent to the PCE providing the affected LSP id, which propagates inter-domain alarms along the BRPC domains chain. Fault localization procedure is then performed at the PCE employing an algorithm based on fuzzy failure set.

VII. PCE IN MULTI-DOMAIN MULTI-CARRIER SCENARIO

Multi-domain multi-provider networks are becoming a promising solution in order to effectively provide world-wide QoS services [144]–[146]. The possibility offered by the PCE architecture to achieve effective TE solutions may be replicated in multi-provider scenarios, in which peer carriers agree on sharing network resources to provide QoS-based services and form a confederated TE-based multi-domain network and adopt service composition strategies to achieve end-to-end QoS. The PCE architecture can be suitable to overcome some of the issues related to inter-provider QoS delivery, such as performance measurements and provider interconnection architectural models [144]. In such scenario, the most suitable architectures are those that enable peer relationship among PCEs, as, for example, per-domain (see Sec. VI-A), standard backward (see Sec. VI-B1) and BRPC (see Sec. VI-B2).

While TE solutions provided by the aforementioned PCE schemes can be suitably applied also in a multi-provider confederated network, additional issues, typical of the inter-carrier scenario, have to be carefully considered. In fact, enhanced inter-operability raises a huge challenge for what concerns path computation policies, security and confidentiality.

Policies are defined as actions in response to network events or conditions based on pre-established rules defined by a network administrator. In [147] a policy-enabled path computation framework is discussed. Inter-carrier path computation can cope with multiple and independent policies established by each network operator. Basic common inter-domain policies should then be agreed among operators within the scope of achieving effective path computation. Moreover, not only TE, but also economical and business reasons drive inter-carrier path computation strategies

In [148] the first inter-provider PCE-based testbed is implemented and evaluated on the basis of the service and management plane architecture proposed in [111]. Moreover, a set of RSVP-TE and PCEP extensions have been proposed to identify and carry the autonomous system chain information, the service identifier and the inter-domain metrics.

In [149] the PCE is adopted to perform cooperative distributed routing optimization with the aim of achieving a fair income distribution. This goal represents one of the most important requirements for the adoption of inter-carrier TE. The authors assume that the income of a domain is proportional to the amount of inter-domain traffic injected in the network and propose an optimization technique derived from cooperative games. Simulation results show that income increases for those domains attracting traffic and allowing high transit traffic volumes, thus confirming that inter-carrier TE has the potential to become one of the reference business models for the internet.

In [150], a business-driven PCE is proposed. The approach is based on the introduction of business metrics, such as SLA matching and pricing, in order to select either the domain sequence and the BRPC end-to-end path. Additional business and QoS database, populated by resorting to economical and SLA policies, is joined to TED in order to select the next peer PCE of the BRPC chain.

In [106], a backward-compatible mechanism is proposed to both enable domain summarization and multipath routing, by encapsulating link disjoint paths onto a single virtual path and by providing multiple virtual intra-domain topologies.

Security and confidentiality should also be preserved. Breaking confidential information is a hot threat to the exchange of data between operators. It can be carried out in many ways, such as intrusion attempts during the data transfer, injection of modified control messages, spoofing, or more sophistically, cross-analysis of the information provided by a domain through the use of an authorized and authenticated protocol session. Typically, to limit the disclosure of confidential information, domains establish and enforce policies assuring a given Service Level Agreement (SLA), able to either guarantee service features and protect the interconnected network together with the associated traffic (e.g., data, control, management traffic). Policies include identification, authentication and authorization mechanisms under the umbrella of a security trust model [151].

Concerning PCEP, one adopted mechanism to preserve confidentiality discourages PCEs and PCCs to exchange strict explicit lists of traversed intra-domain hops and paths are expressed in the form of an encrypted key [152] [109]. However, this basic level of trust agreement may not guarantee the required level of confidentiality. In [100], an overview of security considerations is provided. Requirements and possible solutions are indicated to address vulnerability aspects including spoofing (PCC or PCE impersonation), snooping (message interception), falsification, and denial of service. With reference to confidentiality aspects, [100] identifies the need to additionally define network policies aiming at preserving network information from bogus computation requests. Indeed, differently from connection requests triggered during signaling [153], PCEP-based computations do not imply the subsequent setup of the required connection, thus potentially enabling a malicious utilization of the PCE Architecture. A number of studies have discussed and investigated security and confidentiality in the inter-carrier PCE context.

In [109] the issue of preserving the disclosure of internal network topology information was considered in multi-domain path computation and an implementation of encrypted path computation is evaluated.

Authors of [154], [155] introduced a security mechanism enabling authentication, authorization and accounting functions by integrating path computation and reservation. In particular, the relationship between path computation and reservation is kept secured and authenticated by the exchange of verified path keys, realized with the utilization of secure tokens.

Works [156]–[159] discussed smart mechanisms that allow an authorized and authenticated peer PCE to infer critical intra-domain information of adjacent domains through licit PCEP requests set. This security leak might represent a valuable advantage for a competitor to gain market share leveraging on potential failures and weaknesses in other peer domains. The proposed signature-based [158] and anomaly-based [159] detection mechanisms, along with related policy enforcement mechanisms, are evaluated in terms of detection capability of malicious utilization of PCEP aiming to break confidentiality.

VIII. Conclusion

This survey presented the state-of-the-art and main activity and research efforts carried out in the recent years in the context of the Path Computation Element (PCE) architecture, with specific focus on core networks running the GMPLS control plane. Both packet-switched (e.g., MPLS networks) and lambda-switched (e.g., WSON) networks can experience benefits in terms of overall integrated control plane stability, resource utilization, TE solutions in large multi-layer and multi-domain networks without affecting distributed routing protocols scalability.

The following topics were covered by the survey: the PCE architecture description and motivations, the PCE in single domain, multi-layer, multi-domain and multi-carrier networks, with specific focus on bundling and synchronization, provisioning and protection/restoration techniques, database update strategies and optically impairment-aware computation.

Consolidated results provided by large amount of studies in the literature and thoroughly reported and elaborated in this survey assessed the PCE feasibility, underlying its architectural benefits and most valuable range of applicability. Finally, this survey is intended to stimulate the future scientific community focused on novel control plane solutions. In particular, it aims at contributing to next step in-progress PCE standardization activity efforts carried out within the IETF PCE working group.

Acknowledgment

This work is supported by the FP7 EU STRONGEST (contract number INFSO-ICT 247674) and IDEALIST (contract number INFSO-ICT 317999) projects.

References

[1] E. Mannie, “Generalized Multi-Protocol Label Switching (GMPLS) architecture,” *IETF, RFC 2945*, Oct 2004.

[2] A. Farrel, J. P. Vasseur, and J. Ash, "A Path Computation Element (PCE)-based architecture," *IETF, RFC 4655*, Aug 2006.

[3] "GEYSERS: Generalised architecture for dynamic infrastructure services," http://www.geysers.eu/.

[4] "ETICS: Economics and technologies for inter-carrier services," https://www.ict-etics.eu/.

[5] "STRONGEST: Scalable, tunable and resilient optical networks guaranteeing extremely-high speed transport," http://www.ict-strongest.eu/.

[6] "DRAGON: Dynamic resource allocation via GMPLS optical networks," http://dragon.east.isi.edu.

[7] J. Vasseur and J. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)," *IETF, RFC 5440*, Mar 2009.

[8] J. Le Roux, "Requirements for path computation element (PCE) discovery," *IETF, RFC 4674*, Oct. 2006.

[9] D. O Awduche and B. Jabbari, "Internet traffic engineering using multi-protocol label switching (MPLS)," *Computer Networks*, vol. 40, no. 1, pp. 111 – 129, 2002.

[10] E. Oki, T. Takeda, and A. Farrel, "Extensions to the path computation element communication protocol (PCEP) for route exclusions," *IETF, RFC 5521*, Apr. 2009.

[11] J. Le Roux, J. Vasseur, and L. Y., "Encoding of objective functions in the path computation element communication protocol (PCEP)," *IETF, RFC 5541*, Jun. 2009.

[12] I. Nishioka and D. King, "Use of the synchronized vector (SVEC) list for synchronized dependent path computations," *IETF, RFC 6007*, Oct. 2010.

[13] P. Korus, F. Paolucci, L. Valcarenghi, F. Cugini, P. Castoldi, M. Kantor, and K. Wajda, "Experimental evaluation of batch versus per-request service interconnection activation in PCE-based grid networking," in *High Performance Switching and Routing, 2009. HPSR 2009. International Conference on*, Jun. 2009.

[14] L. Valcarenghi, P. Korus, F. Paolucci, F. Cugini, M. Kantor, K. Wajda, and P. Castoldi, "Experimental evaluation of pce-based batch provisioning of grid service interconnections," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 30 2009-dec. 4 2009.

[15] F. Cugini, S. Xu, H. Harai, F. Paolucci, L. Valcarenghi, and P. Castoldi, "Introducing path computation element (PCE) in optical grid networking," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, Sept. 2008, pp. 1 –2.

[16] ——, "Optical grid networking exploiting path computation element (PCE) architecture," *International Journal of Communication Networks and Distributed Systems*, vol. 5, no. 3, pp. 246 –262, 2010.

[17] J. Ahmed, P. Monti, and L. Wosinska, "Concurrent processing of multiple LSP request bundles on a PCE in a WDM network," in *OFC/NFOEC 2010 Tech. Dig.*, Mar. 2010.

[18] J. Ahmed, C. Cavdar, P. Monti, and L. Wosinska, "A dynamic bulk provisioning framework for concurrent optimization in PCE-based WDM networks," *Lightwave Technology, Journal of*, vol. 30, no. 14, pp. 2229 –2239, Jul.15, 2012.

[19] F. Calderon, W. Creixell, and A. Beghelli, "Buffering-prediction lightpath allocation in PCE-based WDM optical networks," in *Networks and Optical Communications (NOC), 2011 16th European Conference on*, Jul. 2011, pp. 216 –219.

[20] Y. Lee, J. Le Roux, D. King, and E. Oki, "Path computation element communication protocol (PCEP) requirements and protocol extensions in support of global concurrent optimization," *IETF, RFC 5557*, Oct. 2009.

[21] P. Castoldi, F. Paolucci, A. Giorgetti, F. Cugini, L. Valcarenghi, S. Huang, and K.-i. Kitayama, "Management-augmented stateless PCE for wavelength routed optical networks," in *Proc. ICTON 2008*, vol. 3, Jun. 2008, pp. 23 –26.

[22] K. Lu, G. Xiao, and I. Chlamtac, "Analysis of blocking probability for distributed lightpath establishment in wdm optical networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 1, pp. 187 – 197, Feb. 2005.

[23] A. Giorgetti, N. Sambo, I. Cerutti, N. Andriolli, and P. Castoldi, "Label preference schemes for lightpath provisioning and restoration in distributed gmpls networks," *Lightwave Technology, Journal of*, vol. 27, no. 6, pp. 688–697, Mar. 2009.

[24] O. de Dios, F. Chico, and F. del Nuevo, "Benefits of limited context awareness in stateless PCE," in *OFC/NFOEC 2011 Tech. Dig.*, Mar. 2011.

[25] E. Crabbe, J. Medved, and R. Varga, "PCEP extensions for stateful PCE, draft-crabbe-pce-stateful-pce-01," *IETF, PCE WG*, Oct. 2011.

[26] F. Zhang, X. Zhang, Y. Lee, R. Casellas, and O. Gonzalez de Dios, "draft-zhang-pce-stateful-pce-app-00," *IETF, PCE WG*, Mar. 2012.

[27] E. Danna, S. Mandal, and A. Singh, "A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering," in *Proc. INFOCOM 2012*, Mar. 2012, pp. 846 –854.

[28] R. Munoz, R. Casellas, and R. Martinez, "An open GMPLS-enabled control plane testbed for remote development and experimentation of PCE-based path computation algorithms," in *Future Network Mobile Summit (FutureNetw), 2011*, Jun. 2011.

[29] M. Chamania, M. Drogon, and A. Jukan, "Lessons learned from implementing a path computation element (PCE) emulator," in *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference*, Mar. 2011.

[30] ——, "An open-source path computation element (PCE) emulator: Design, implementation, and performance," *Lightwave Technology, Journal of*, vol. 30, no. 4, pp. 414 –426, Feb.15, 2012.

[31] E. Grampin, A. Castro, M. German, F. Rodriguez, G. Tejera, and M. Sanguinetti, "A PCE-based connectivity provisioning management framework," in *Network Operations and Management Symposium, 2007. LANOMS 2007. Latin American*, Sept. 2007, pp. 76 –83.

[32] 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," *IEEE Commun. Lett.*, vol. 14, no. 6, pp. 575–577, Jun. 2010.

[33] D. Alvarez, V. Lopez, J. Anamuro, J. de Vergara, O. de Dios, and J. Aracil, "Utilization of temporary reservation of path computed resources for multi-domain path computation element protocols in WDM networks," in *Network of the Future (NOF), 2011 International Conference on the*, Nov. 2011, pp. 102 –106.

[34] Y. Lee, G. Bernstein, J. Martensson, T. Takeda, T. Tsuritani, and O. de Dios, "PCEP requirements for WSON routing and wavelength assignment, draft-ietf-pce-wson-routing-wavelength-06," *IETF, PCE WG*, Oct. 2011.

[35] A. Giorgetti, F. Paolucci, F. Cugini, L. Valcarenghi, P. Castoldi, and G. Bernstein, "Routing and wavelength assignment in PCE-based wavelength switched optical networks," in *Proc. ECOC 2008*, Sept. 2008.

[36] Y. Zhao, J. Zhang, H. Huang, W. Gu, and Y. Ji, "Analysis of PCE-based wavelength assignment schemes in wavelength switched optical networks," in *Wireless and Optical Communications Conference (WOCC), 2010 19th Annual*, May 2010.

[37] 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, Apr. 2010.

[38] A. Giorgetti, N. Andriolli, S. Ruepp, and P. Castoldi, "PCE-based vs. distributed set up of bidirectional lightpaths in GMPLS-controlled optical networks," in *OFC/NFOEC 2010 Tech. Dig.*, Mar. 2010.

[39] L. Hou, L. Zuo, P. Luo, and D. Liu, "Collision-safe wavelength assignment scheme for lightpath establishment in WSONs," in *Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on*, Sept. 2011, pp. 250 –253.

[40] Y. Pointurier, M. Brandt-Pearce, and S. Subramaniam, "Analysis of blocking probability in noise- and cross-talk-impaired all-optical networks," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 1, no. 6, pp. 543 –554, Nov. 2009.

[41] F. Cugini, N. Sambo, N. Andriolli, A. Giorgetti, L. Valcarenghi, P. Castoldi, E. Le Rouzic, and J. Poirrier, "Enhancing GMPLS signaling protocol for encompassing quality of transmission (QoT) in all-optical networks," *IEEE J. Lightw. Technol.*, vol. 26, no. 19, pp. 3318 –3328, Oct. 2008.

[42] A. Bononi, M. Bertolini, P. Serena, and G. Bellotti, "Cross-phase modulation induced by OOK channels on higher-rate DQPSK and coherent QPSK channels," *IEEE J. Lightw. Technol.*, vol. 27, no. 18, pp. 3974 –3983, Sept.15, 2009.

[43] N. Sambo, M. Secondini, F. Cugini, G. Bottari, P. Iovanna, F. Cavaliere, and P. Castoldi, "Modeling and distributed provisioning in 10-40-100-gb/s multirate wavelength switched optical networks," *IEEE J. Lightw. Technol.*, vol. 29, no. 9, pp. 1248 –1257, May 2011.

[44] N. Sambo, F. Cugini, I. Cerutti, L. Valcarenghi, P. Castoldi, J. Poirrier, E. Le Rouzic, and C. Pinart, "Probe-based schemes to guarantee lightpath quality of transmission (QoT) in transparent optical networks," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, Sept. 2008.

[45] N. Sambo, P. Castoldi, F. Cugini, G. Bottari, and P. Iovanna, "Toward high-rate and flexible optical networks," *IEEE Commun. Mag.*, vol. 50, no. 5, pp. 66 –72, May 2012.

[46] Y. Lee, G. Bernstein, J. Martensson, T. Takeda, and T. Tsuritani, "PCEP extensions for WSON impairments, draft-lee-pce-wson-impairments-03.txt," *IETF*, Jan. 2012.

[47] F. Cugini, F. Paolucci, L. Valcarenghi, and P. Castoldi, "Implementing a path computation element (PCE) to encompass physical impairments in transparent networks," in *OFC/NFOEC 2007 Tech. Dig.*, Mar. 2007.

[48] T. Tsuritani, M. Miyazawa, S. Kashihara, and T. Otani, "Optical path computation element interworking with network management system for transparent mesh networks," in *OFC/NFOEC 2008 Tech. Dig.*, Feb. 2008.

[49] F. Agraz, S. Azodolmolky, M. Angelou, J. Perello, L. Velasco, S. Spadaro, A. Francescon, C. Saradhi, Y. Pointurier, P. Kokkinos, E. Varvarigos, M. Gunkel, and I. Tomkos, "Experimental demonstration of centralized and distributed impairment-aware control plane schemes for dynamic transparent optical networks," in *OFC/NFOEC 2010 Tech. Dig.*, Mar. 2010.

[50] R. Martinez, R. Casellas, R. Munoz, and T. Tsuritani, "Experimental study and assessment of the link distance impact in an OSNR-based IRWA algorithm in GMPLS-enabled translucent WSON networks," in *Proc. ICTON 2010*, Jul. 2010.

[51] S. Spadaro, J. Perello, F. Agraz, S. Azodolmolky, M. Angelou, Y. Qin, R. Nejabati, D. Simeonidou, P. Kokkinos, E. Varvarigos, Y. Ye, and I. Tomkos, "Experimental demonstration of an enhanced impairment-aware path computation element," in *OFC/NFOEC 2011 Tech. Dig.*, Mar. 2011.

[52] Y. Wang, J. Zhang, Y. Zhao, X. Cao, H. Huang, W. Gu, and Y. Ji, "Double cooperating PCEs for IA-RWA computation with QoT compensation in transparent optical networks," in *Proc. ECOC 2010*, Sept. 2010.

[53] J. H. Lee, T. Tsuritani, H. Guo, S. Okamoto, N. Yoshikane, and T. Otani, "Field trial of GMPLS-controlled all-optical networking assisted with optical performance monitors," in *OFC/NFOEC 2008 Tech. Dig.*, Feb. 2008.

[54] M. Suzuki and T. Tsuritani, "Management and control of transparent optical network considering physical impairments," in *Transparent Optical Networks, 2009. ICTON '09. 11th International Conference on*, Jul. 2009.

[55] D. Caviglia, G. Bottari, and F. Lazzeri, "Photonic network optimization via resource and LSPs concurrent allocation: A PCE concept," in *OFC/NFOEC 2010 Tech. Dig.*, Mar. 2010.

[56] N. Sambo, F. Cugini, M. Secondini, P. Iovanna, G. Bottari, and P. Castoldi, "Enforcing QoT via PCE in multi bit-rate WSONs," *Communications Letters, IEEE*, vol. 15, no. 4, pp. 452 –454, Apr. 2011.

[57] F. Paolucci, N. Sambo, F. Cugini, A. Giorgetti, and P. Castoldi, "Experimental demonstration of impairment-aware PCE for multi-bit-rate WSONs," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 3, no. 8, pp. 610 –619, Aug. 2011.

[58] N. Sambo, F. Paolucci, G. Meloni, F. Cugini, M. Secondini, L. Poti, and P. Castoldi, "Candidate paths for impairment-aware PCE in 10-100 gb/s optical networks," *Communications Letters, IEEE*, vol. 15, no. 8, pp. 881 –883, Aug. 2011.

[59] M. Jinno, T. Ohara, Y. Sone, A. Hirano, O. Ishida, and M. Tomizawa, "Elastic and adaptive optical networks: possible adoption scenarios and future standardization aspects," *IEEE Commun. Mag.*, vol. 49, no. 10, pp. 164 –172, Oct. 2011.

[60] O. Gerstel, M. Jinno, A. Lord, and S. Yoo, "Elastic optical networking: a new dawn for the optical layer?" *IEEE Commun. Mag.*, vol. 50, no. 2, pp. s12 –s20, Feb. 2012.

[61] F. Zhang, X. Zi, O. Gonzales de Dios, and R. Casellas, "Requirements for GMPLS control of flexible grids, draft-zhang-ccamp-flexible-grid-requirements-01.txt," *IETF, CCAMP WG*, Oct. 2011.

[62] G. Meloni, F. Paolucci, N. Sambo, F. Cugini, M. Secondini, L. Gerardi, L. Poti, and P. Castoldi, "PCE architecture for flexible WSON enabling dynamic rerouting with modulation format adaptation," in *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on*, Sept. 2011.

[63] F. Cugini, G. Meloni, F. Paolucci, N. Sambo, M. Secondini, L. Gerardi, L. Poti, and P. Castoldi, "Demonstration of flexible optical network based on path computation element," *IEEE J. Lightw. Technol.*, vol. 30, no. 5, pp. 727 –733, Mar. 2012.

[64] R. Casellas, R. Munoz, J. Fabrega, M. Svaluto Moreolo, R. Martinez, L. Liu, T. Tsuritani, and I. Morita, "Experimental assessment of a combined PCE-RMA and distributed spectrum allocation mechanism for GMPLS elastic CO-OFDM optical networks," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2012 Conference on (OFC/NFOEC)*, Mar. 2012.

[65] A. Patel, P. Ji, J. Jue, and T. Wang, "Defragmentation of transparent flexible optical WDM (FWDM) networks," in *Optical Fiber Communication Conference (OFC/NFOEC)*, Mar. 2011.

[66] T. Takagi, H. Hasegawa, K. Sato, Y. Sone, A. Hirano, and M. Jinno, "Disruption minimized spectrum defragmentation in elastic optical path networks that adopt distance adaptive modulation," in *Proc. ECOC*, Sept. 2011.

[67] K. Wen, Y. Yin, D. Geisler, S. Chang, and B. Yoo, "Dynamic on-demand lightpath provisioning using spectral defragmentation in flexible bandwidth networks," in *Proc. ECOC*, Sept. 2011.

[68] F. Cugini, F. Paolucci, G. Meloni, G. Berrettini, M. Secondini, F. Fresi, N. Sambo, L. Poti, and P. Castoldi, "Push-pull defragmentation without traffic disruption in flexible grid optical networks," *Lightwave Technology, Journal of*, vol. 31, no. 1, pp. 125 –133, jan.1, 2013.

[69] Y. Zhao, J. Zhang, T. Peng, X. Yu, X. Cao, and D. Wang, "PCEP protocol extension for spectrum utilization optimization in flexi-grid networks, draft-zhaoyl-pce-flexi-grid-pcep-ex-00," *IETF, PCE WG*, Oct. 2011.

[70] G. Li, D. Wang, C. Kalmanek, and R. Doverspike, "Efficient distributed restoration path selection for shared mesh restoration," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 761 – 771, Oct. 2003.

[71] N. Andriolli, A. Giorgetti, L. Valcarenghi, and P. Castoldi, "Idle protection capacity reuse in multiclass optical networks," *IEEE J. Lightw. Technol.*, vol. 25, no. 5, pp. 1152 –1162, May 2007.

[72] M. Tornatore, F. De Grandi, R. Munoz, R. Martinez, R. Casellas, and A. Pattavina, "Effects of outdated control information in control-plane-enabled optical networks with path protection," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 1, no. 2, pp. A194 –A204, Jul. 2009.

[73] F. Kolberg and D. Harle, "Using resource pools for pre-signaled restoration LSP in ASON networks," in *Communications, 2007. ICC '07. IEEE International Conference on*, Jun. 2007, pp. 2242 –2248.

[74] R. Casellas, R. Munoz, and R. Martinez, "A path computation element for shared path protection in GMPLS-enabled wavelength switched optical networks," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, Sept. 2008.

[75] J. Ahmed, C. Cavdar, P. Monti, and L. Wosinska, "Bulk provisioning of LSP requests with shared path protection in a PCE-based WDM network," in *Optical Network Design and Modeling (ONDM), 2011 15th International Conference on*, Feb. 2011.

[76] A. Giorgetti, N. Sambo, I. Cerutti, N. Andriolli, and P. Castoldi, "Label preference schemes for lightpath provisioning and restoration in distributed GMPLS networks," *IEEE J. Lightw. Technol.*, vol. 27, no. 6, pp. 688 –697, Mar15. 2009.

[77] M. Abouelela and M. El-Darieby, "PCE-based hierarchical segment restoration," in *Integrated Network Management, 2009. IM '09. IFIP/IEEE International Symposium on*, Jun. 2009, pp. 164 –171.

[78] A. Giorgetti, L. Valcarenghi, F. Cugini, and P. Castoldi, "PCE-based dynamic restoration in wavelength switched optical networks," in *Proc. ICC 2010*, May 2010.

[79] R. Munoz, R. Casellas, and R. Martinez, "Experimental evaluation of dynamic PCE-based path restoration with centralized and distributed wavelength assignment in GMPLS-enabled transparent wson networks," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, Sept. 2010.

[80] L. Liu, T. Tsuritani, R. Casellas, and M. Tsurusawa, "Demonstration of a resilient PCE/GMPLS controlled translucent optical network," in *OptoeElectronics and Communications Conference (OECC), 2010 15th*, Jul. 2010, pp. 88 –89.

[81] K. Shiomoto, D. Papadimitriou, J. Le Roux, M. Vigoureux, and D. Brungard, "Requirements for GMPLS-based Multi-Region and Multi-Layer Networks (MRN/MLN)," July 2008.

[82] J. Le Roux and D. Papadimitriou, "Evaluation of existing GMPLS protocols against multi-layer and multi-region networks (MLN/MRN)," Sept. 2008.

[83] E. Oki, D. Takeda, J. Le Roux, and A. Farrel, "Framework for PCE-based inter-layer MPLS and GMPLS Traffic Engineering," Sept. 2009.

[84] F. Cugini, A. Giorgetti, N. Andriolli, I. Paolucci, L. Valcarenghi, and P. Castoldi, "Multiple path computation element (PCE) cooperation for multi-layer traffic engineering," in *OFC/NFOEC 2007 Tech. Dig.*, Mar. 2007.

[85] E. Oki, I. Inoue, and K. Shiomoto, "Path computation element (PCE)-based traffic engineering in MPLS and GMPLS networks," in *Sarnoff Symposium, 2007 IEEE*, May 2007.

[86] T. Miyamura, T. Kurimoto, R. Hayashi, I. Inoue, K. Shiomoto, and S. Urushidani, "Demonstration of PCE-controlled dynamic traffic engineering for GMPLS-based multilayer service network," in *Optical Communication, 2005. ECOC 2005. 31st European Conference on*, vol. 2, Sept. 2005, pp. 269 – 270 vol.2.

[87] E. Modiano, "Traffic grooming in WDM networks," *Communications Magazine, IEEE*, vol. 39, no. 7, pp. 124 –129, Jul. 2001.

[88] K. Zhu and B. Mukherjee, "Traffic grooming in an optical WDM mesh network," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 1, pp. 122 –133, Jan. 2002.

[89] F. Agraz, L. Velasco, J. Perello, M. Ruiz, S. Spadaro, G. Junyent, and J. Comellas, "Design and implementation of a GMPLS-controlled grooming-capable optical transport network," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 1, no. 2, pp. A258 –A269, July 2009.

[90] N. Andriolli, F. Cugini, L. Valcarenghi, P. Castoldi, and A. Welin, "Virtual network topology manager (VNTM) and path computation element (PCE) cooperation in multi-layer GMPLS networks," in *OFC/NFOEC 2009 Tech. Dig.*, Mar. 2009.

[91] S. Gunreben and F. Rambach, "Assessment and performance evaluation of PCE-based inter-layer traffic engineering," in *Optical Network Design and Modeling, 2008. ONDM 2008. International Conference on*, Mar. 2008.

[92] M. Chamania, X. Chen, A. Jukan, F. Rambach, C. Gruber, and M. Hoffmann, "Embedding optical ethernet services within the path computation element framework: The 100GET approach," in *OFC/NFOEC 2009 Tech. Dig.*, Mar 2009.

[93] D. Katz, K. Kompella, and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2," *IETF, RFC 3630*, Sep 2003.

[94] F. Ricciato, U. Monaco, and D. Ali, "Distributed schemes for diverse path computation in multidomain MPLS networks," *IEEE Commun. Mag.*, vol. 43, no. 6, pp. 138 – 146, Jun. 2005.

[95] S. Secci, J. Rougier, and A. Pattavina, "AS-level source routing for multi-provider connection-oriented services," *Computer Networks"*, vol. 54, no. 14, pp. 2453 – 2467, 2010.

[96] N. Bithar, R. Zhang, and K. Kumaki, "Inter-AS requirements for the Path Computation Element Communication Protocol (PCECP)," *IETF, RFC 5376*, Nov. 2008.

[97] P. Torab, B. Jabbari, Q. Xu, S. Gong, X. Yang, T. Lehman, C. Tracy, and J. Sobieski, "On cooperative inter-domain path computation," in *Proc. ISCC 06*, Jun. 2006, pp. 511 – 518.

[98] J. Vasseur, A. Ayyangar, and R. Zhang, "A per-domain path computation method for establishing inter-domain traffic engineering (TE) label switched paths (LSPs)," *IETF, RFC 5152*, Feb. 2008.

[99] T. Saad, J. Israr, S. Sivabalan, and H. Mouftah, "An evaluation for PCE selection schemes for inter-domain path computation," in *Transparent Optical Networks, 2007. ICTON '07. 9th International Conference on*, vol. 3, Jul. 2007, p. 187.

[100] J. Vasseur, R. Zhang, N. Bitar, and J. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths," *IETF, RFC 5441*, Apr 2009.

[101] S. Dasgupta, J. 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, Jul-Aug. 2007.

[102] T. Takeda, R. Sugiyama, E. Oki, I. Inoue, K. Shiomoto, K. Shindome, K. Fujihara, and S.-I. Kato, "Implementation and experiments of path computation element-based inter-domain network control and management," in *Proc. ECOC 2008*, Sept. 2008.

[103] G. Geleji, H. Perros, Y. Xin, and T. Beyene, "A performance analysis of inter-domain QoS routing schemes based on path computation elements," in *Proc. HONET 2008*, Nov. 2008, pp. 146 –152.

[104] G. Geleji and H. Perros, "QoS routing across multiple autonomous systems using the path computation element architecture," *Annals of Telecommunications*, vol. 66, pp. 293–306, 2011. [Online]. Available: http://dx.doi.org/10.1007/s12243-010-0206-y

[105] J. Yu, Y. He, K. Wu, M. Tacca, A. Fumagalli, and J.-P. Vasseur, "A queueing model framework of PCE-based inter-area path computation," in *Proc. INFOCOM 2009*, Apr. 2009, pp. 2666 –2670.

[106] X. Chen, M. Chamania, and A. Jukan, "A backward-compatible inter-domain multipath routing framework," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, Apr. 2011, pp. 133 –138.

[107] X. Wan, Y. Chen, H. Zhang, and X. Zheng, "Dynamic domain-sequencing scheme for inter-domain path computation in WDM networks," in *Proc. ACP 2009*, vol. 2009-Supplement, Nov. 2009.

[108] J. Li, J. Zhang, Y. Zhao, W. Gu, and Y. Ji, "Impairment-aware backward recursive PCE-based computation algorithm for wavelength switched optical networks," in *Communication Technology (ICCT), 2010 12th IEEE International Conference on*, Nov. 2010, pp. 693 –696.

[109] R. Casellas, R. Martinez, R. Munoz, T. Tsuritani, L. Liu, and M. Tsurusawa, "Lab-trial of multi-domain lightpath provisioning with PCE path computation combining BRPC and path-key topology confidentiality in GMPLS translucent WSON networks," in *Proc. ECOC 2010*, Sept. 2010.

[110] R. Casellas, R. Martinez, R. Munoz, and S. Gunreben, "Enhanced backwards recursive path computation for multi-area wavelength switched optical networks under wavelength continuity constraint," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 1, no. 2, pp. A180 –A193, 2009.

[111] R. Douville, J.-L. L. Roux, J.-L. Rougier, and S. Secci, "A service plane over the PCE architecture for automatic multidomain connection-oriented services," *IEEE Commun. Mag.*, vol. 46, no. 6, pp. 94–102, 2008.

[112] S. Secci, J.-L. Rougier, and A. Pattavina, "AS tree selection for inter-domain multipoint MPLS tunnels," in *Communications, 2008. ICC '08. IEEE International Conference on*, May 2008, pp. 5863 –5868.

[113] "External Network-Network Interface (E-NNI) OSPF-based routing - 1.0 (Intra-Carrier) implementation agreement," *Optical Internetworking Forum, OIF*, 2007.

[114] F. Paolucci, F. Cugini, A. Giorgetti, L. Valcarenghi, P. Castoldi, P. Iovanna, G. Bottari, and A. Welin, "PCE architecture for OIF E-NNI multi-domain routing evaluated in an intra-domain WSON scenario," in *Optical Communication, 2009. ECOC '09. 35th European Conference on*, Sept. 2009.

[115] F. Paolucci, F. Cugini, P. Iovanna, G. Bottari, L. Valcarenghi, and P. Castoldi, "Delay-bandwidth-aware metric abstraction schemes for OIF E-NNI multidomain traffic engineering," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 2, no. 10, pp. 782 –792, Oct. 2010.

[116] S. Spadaro, J. Perello and, G. Hernandez-Sola, A. Moreno, F. Agraz, J. Comellas, and G. Junyent, "Analysis of traffic engineering information dissemination strategies in PCE-based multi-domain optical networks," in *Proc. ICTON 2010*, Jul. 2010.

[117] A. Pages, R. Casellas, J. Perello, R. Martinez, S. Spadaro, and R. Munoz, "Experimental evaluation of a full-meshed domain abstraction design model for reduced state information dissemination in multi-domain PCE-based wsons," in *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on*, Sept. 2011.

[118] F. Cugini, F. Paolucci, L. Valcarenghi, P. Castoldi, and A. Welin, "PCE communication protocol for resource advertisement in multi-domain BGP-based networks," *Optical Fiber Communication 2009 Tech. Dig.*, Mar. 2009.

[119] L. Bisti, E. Mingozzi, and G. Stea, "Interdomain path computation for PCE-assisted traffic engineering," in *Proc. BROADNETS 2009*, Sept. 2009.

[120] L. Buzzi, M. Bardellini, D. Siracusa, G. Maier, F. Paolucci, F. Cugini, L. Valcarenghi, and P. Castoldi, "Hierarchical border gateway protocol (HBGP) for PCE-based multi-domain traffic engineering," in *Proc. ICC 2010*, May 2010.

[121] A. Pontes, A. Drummond, N. Da Fonseca, and A. Jukan, "PCE-based inter-domain lightpath provisioning," in *Proc. ICC 2012*, Jun. 2012.

[122] D. Siracusa, S. Grita, G. Maier, A. Pattavina, F. Paolucci, F. Cugini, and P. Castoldi, "Domain sequence protocol (DSP) for PCE-based multi-domain traffic engineering," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 4, no. 11, pp. 876 –884, nov. 2012.

[123] D. King and A. Farrel, "draft-ietf-pce-hierarchy-fwk-00.txt," *IETF, PCE WG*, Oct. 2011.

[124] R. Casellas, R. Munoz, and R. Martinez, "Lab trial of multi-domain path computation in GMPLS controlled WSON using a hierarchical PCE," in *OFC/NFOEC 2011 Tech. Dig.*, Mar. 2011.

[125] R. Casellas, R. Martinez, R. Munoz, L. Liu, T. Tsuritani, I. Morita, and M. Tsurusawa, "Dynamic virtual link mesh topology aggregation in multi-domain translucent WSON with hierarchical-PCE," in *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on*, Sept. 2011.

[126] 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 (OFC), collocated National Fiber Optic Engineers Conference, 2012 Conference on (OFC/NFOEC)*, Mar. 2012.

[127] H. Matsuura, N. Morita, T. Murakami, and K. Takami, "Hierarchically distributed PCS for GMPLS multilayered networks," in *Proc. GLOBECOM 2005.*, vol. 4, Dec. 2005, pp. 6 pp. –2077.

[128] S. Shang, X. Zheng, H. Zhang, N. Hua, and H. Zhang, "A hierarchical path computation element (PCE)-based routing algorithm in multi-domain WDM networks," in *Proc. ACP 2010*, Dec. 2010, pp. 35 –36.

[129] G. Hernandez-Sola, J. Perello, F. Agraz, S. Spadaro, J. Comellas, and G. Junyent, "Scalable hybrid path computation procedure for PCE-based multi-domain WSON networks," in *Transparent Optical Networks (ICTON), 2011 13th International Conference on*, Jun. 2011.

[130] A. Giorgetti, F. Paolucci, F. Cugini, and P. Castoldi, "Hierarchical PCE in GMPLS-based multi-domain wavelength switched optical networks," in *Optical Fiber Communication Conference and Exposi-*

tion (OFC/NFOEC), 2011 and the National Fiber Optic Engineers Conference, Mar. 2011.

[131] ——, "Impact of intra-domain information in GMPLS-based WSONs with hierarchical PCE," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2012 Conference on (OFC/NFOEC)*, Mar. 2012.

[132] M. Chamania, X. Chen, A. Jukan, F. Rambach, C. Gruber, and M. Hoffmann, "Adaptive advance reservation based inter-domain framework," in *Advanced Networks and Telecommunication Systems, 2008. ANTS '08. 2nd International Symposium on*, Dec. 2008.

[133] H. Guo, T. Tsuritani, S. Okamoto, and T. Otani, "Demonstration of GMPLS-controlled inter-domain transparent optical networks," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, Sept. 2008.

[134] L. Liu, T. Tsuritani, R. Casellas, R. Martinez, R. Munoz, and M. Tsurusawa, "Experimental demonstration and comparison of distributed and centralized multi-domain resilient translucent WSON," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, Sept. 2010.

[135] R. Martinez, R. Casellas, L. Velasco, F. Agraz, R. Munoz, and S. Spadaro, "Experimental validation of end-to-end GMPLS-enabled restoration in multi-domain transparent WSON," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, Mar. 2010.

[136] L. Velasco, F. Agraz, R. Martinez, R. Casellas, S. Spadaro, R. Munoz, and G. Junyent, "GMPLS-based multidomain restoration: Analysis, strategies, policies and experimental assessment," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 2, no. 7, pp. 427 –441, Jul. 2010.

[137] F. Paolucci, F. Cugini, L. Valcarenghi, and P. Castoldi, "Enhancing backward recursive PCE-based computation (BRPC) for inter-domain protected LSP provisioning," in *Optical Fiber communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on*, Feb. 2008.

[138] G. Hernndez-Sola, J. Perell, F. Agraz, L. Velasco, S. Spadaro, and G. Junyent, "Enhanced domain disjoint backward recursive TE path computation for PCE-based multi-domain networks," *Photonic Network Communications*, vol. 21, pp. 141–151, 2011, 10.1007/s11107-010-0288-5. [Online]. Available: http://dx.doi.org/10.1007/s11107-010-0288-5

[139] Z. Chen, Y. Peng, Z. Wang, Y. Wang, X. Chen, and K. Long, "An enhanced backward recursive PCE-based computation scheme for end-to-end disjoint paths in multi-domain networks," *Photonic Network Communications*, pp. 1–7, 10.1007/s11107-011-0363-6. [Online]. Available: http://dx.doi.org/10.1007/s11107-011-0363-6

[140] Q. Zhang, M. Hasan, X. Wang, P. Palacharla, and M. Sekiya, "Efficient PCE-based survivable path computation in multi-domain networks," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, Apr. 2011, pp. 139 –144.

[141] X. Cao, J. Zhang, Y. Zhao, J. Liu, D. Han, and W. Gu, "A PCE-based fast reroute algorithm for multi-failures in multi-domain optical networks," in *Proc. ACP 2009*, Nov. 2009.

[142] L. Liu, R. Casellas, T. Tsuritani, I. Morita, R. Martinez, and R. Mu andoz, "Lab trial of PCE-based OSNR-aware dynamic restoration in multi-domain GMPLS-enabled translucent WSON," in *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on*, Sept. 2011.

[143] J. Luo, S. Huang, J. Zhang, X. Li, and W. Gu, "A novel multi-fault localization mechanism in PCE-based multi-domain large capacity optical transport networks," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2012 Conference on (OFC/NFOEC)*, Mar. 2012.

[144] P. Jacob and B. Davie, "Technical challenges in the delivery of interprovider QoS," *IEEE Commun. Mag.*, vol. 43, no. 6, pp. 112 – 118, Jun. 2005.

[145] L. Fang, N. Bita, J.-L. Le Roux, and J. Miles, "Interprovider IP-MPLS services: requirements, implementations, and challenges," *IEEE Commun. Mag.*, vol. 43, no. 6, pp. 119 – 128, Jun. 2005.

[146] M. Yampolskiy, W. Hommel, B. Lichtinger, W. Fritz, and M. Hamm, "Multi-domain end-to-end (E2E) routing with multiple QoS parameters - considering real world user requirements and service provider constraints," in *Evolving Internet (INTERNET), 2010 Second International Conference on*, Sept. 2010, pp. 9 –18.

[147] I. Bryskin, D. Papadimitriou, L. Berger, and J. Ash, "Policy-enabled path computation framework," *IETF, RFC 5394*, Dec. 2008.

[148] A. Bianzino, J.-L. Rougier, S. Secci, R. Casellas, R. Martinez, R. Munoz, N. Djarallah, R. Douville, and H. Pouyllau, "Testbed implementation of control plane extensions for inter-carrier GMPLS LSP provisioning," in *Testbeds and Research Infrastructures for the Development of Networks Communities and Workshops, 2009. TridentCom 2009. 5th International Conference on*, Apr. 2009.

[149] M. Mycek, S. Secci, M. Pioro, J.-L. Rougier, A. Tomaszewski, and A. Pattavina, "Cooperative multi-provider routing optimization and income distribution," in *Design of Reliable Communication Networks, 2009. DRCN 2009. 7th International Workshop on*, Oct. 2009, pp. 281 –288.

[150] N. Djarallah, H. Pouyllau, N. Le Sauze, and R. Douville, "Business-driven PCE for inter-carrier QoS connectivity services," in *Future Network Mobile Summit (FutureNetw), 2011*, Jun. 2011.

[151] "Security mechanisms and procedures for NGN," *ITU-T Recommendation Y.2704.2010*, Jan. 2010.

[152] R. Bradford, J.-P. Vasseur, and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Key-Based Mechanism," *IETF, RFC 5520*, Apr. 2009.

[153] E. Toktar, E. Jamhour, and E. Maziero, "RSVP policy control using XACML," in *Policies for Distributed Systems and Networks, 2004. POLICY 2004. Proceedings. Fifth IEEE International Workshop on*, Jun. 2004, pp. 87 – 96.

[154] S. Polito, M. Chamania, and A. Jukan, "Extending the inter-domain PCE framework for authentication and authorization in GMPLS networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, Jun. 2009.

[155] S. Polito, S. Zaghloul, M. Chamania, and A. Jukan, "Inter-domain path provisioning with security features: Architecture and signaling performance," *IEEE Trans. Network Service Management*, vol. 8, no. 3, pp. 219 –233, Sept. 2011.

[156] F. Paolucci, F. Cugini, B. Martini, M. Gharbaoui, L. Valcarenghi, and P. Castoldi, "Preserving confidentiality in PCEP-based inter-domain path computation," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, Sept. 2010.

[157] M. Gharbaoui, F. Paolucci, B. Martini, F. Cugini, and P. Castoldi, "Behavior-based policies for preserving confidentiality in PCE-based multi-domain MPLS networks," in *Policies for Distributed Systems and Networks (POLICY), 2011 IEEE International Symposium on*, Jun. 2011, pp. 81 –88.

[158] F. Paolucci, M. Gharbaoui, A. Giorgetti, F. Cugini, B. Martini, L. Valcarenghi, and P. Castoldi, "Preserving confidentiality in PCE-based multi-domain networks," *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 3, no. 5, pp. 465 –474, May 2011.

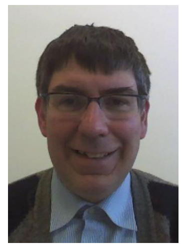
[159] M. Gharbaoui, F. Paolucci, A. Giorgetti, B. Martini, and P. Castoldi, "Statistical approach for detecting malicious PCE activity in multi-domain networks," in *Proc. HPSR 2012*, Jun. 2012.

Francesco Paolucci received the Laurea degree in Telecommunications Engineering in 2002 from the University of Pisa, Italy and the Ph.D. degree in 2009 from the Scuola Superiore Sant'Anna, Pisa, Italy. In 2008 he was granted a research Merit Scholarship at the Istitut National de le Recherche Scientifique (INRS), Montreal, Quebec, Canada. Currently, he is Postdoctoral Research Fellow at the TeCIP Institute of Scuola Superiore Sant'Anna, Pisa, Italy. His main research interests are in the field of optical networks control plane, including Generalized Multi Protocol Label Switching (GMPLS) and Software Defined Networking (SDN) protocol extensions, impairment-aware routing based on Path Computation Element (PCE), inter-domain traffic engineering and flexible optical node architectures. Other research activities include network services for Grid/Cloud Computing applications, optical network fault tolerance, inter-domain security and confidentiality.

Filippo Cugini received the M.S. degree in Telecommunication Engineering from the University of Parma, Italy. Since 2001, he has been with the National Lab. of Photonic Networks, Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Pisa, Italy. He is co-author of ten international patents and more than eighty IEEE publications.

Alessio Giorgetti received the Laurea degree in Telecommunications Engineering in 2002 from the University of Pisa, Italy, and the Ph.D. degree in 2006 from Scuola Superiore Sant'Anna (SSSUP), Pisa, Italy. From December 2006 to September 2007 he was a Visiting Scholar at the Centre for Advanced Photonics and Electronics (CAPE), Cambridge, UK. From October 2007 to June 2011 he was a post-doc at SSSUP. Since July 2011 he is an Assistant Professor at SSSUP. His main research interests include design and planning of optical networks design, fault tolerance, traffic grooming, GMPLS control plane, next generation SONET/SDH technologies, WDM-PON access networks design, Industrial Ethernet networks.

Nicola Sambo received his Laurea degree cum laude in telecommunication engineering from the University of Pisa, Italy, in 2004 and his Ph.D. degree from Scuola Superiore Sant'Anna, Pisa, Italy, in 2009. Currently, he is Assistant Professor at Scuola Superiore Sant'Anna. He is also collaborating with CNIT, Pisa, Italy and Ericsson, Pisa, Italy. His research interests include WDM network planning and control plane, transmission modeling, coherent transmission, GMPLS protocol suite and PCE, routing and set up.



Piero Castoldi (IEEE Senior Member) has been Associate Professor in Telecommunications and Responsible for the "Networks and Services" area at Scuola Superiore Sant'Anna (Pisa, Italy) since 2001, as well as coordinator of the Communication macro-area at the TeCIP Institute since 2012. He holds a Master degree in Electrical Engineering from the University of Bologna (Italy) and a PhD degree in Information Technology from the University of Parma (Italy). He has also been visiting Post-doc and Researcher at the Department of Electrical Engineering at Princeton University (USA) for a total of 18 months in the period 1996-2000. He also collaborates with the Italian Inter-universitary Consortium for Telecommunications (CNIT) and he has been Director of the National Laboratory for Photonic Networks in the period 2005-2010. His scientific activity has covered the area of digital transmission, telecommunication networks performance, and, more recently, architecture, protocols and switching techniques for optical networks and optical interconnection networks. He is author of more than 200 technical papers published in international conferences and journals and single-author of a book on CDMA published by Artech House. He has also filed a total of 12 patents.