PCE Working Group
Internet Draft
Intended status: Informational
Expires: December 3, 2014

D. King
Old Dog Consulting
J. Meuric
O. Dugeon
France Telecom
Q. Zhao
Huawei Technologies
Oscar Gonzalez de Dios
Telefonica I+D

June 3, 2014

Applicability of the Path Computation Element to Inter-Area and Inter-AS MPLS and GMPLS Traffic Engineering

draft-ietf-pce-inter-area-as-applicability-04

Abstract

The Path Computation Element (PCE) may be used for computing services that traverse multi-area and multi-AS Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineered (TE) networks.

This document examines the applicability of the PCE architecture, protocols, and protocol extensions for computing multi-area and multi-AS paths in MPLS and GMPLS networks.

Copyright Notice

Table of Contents

1. Introduction..3
  1.1. Domains..4
  1.2. Path Computation...4
    1.2.1 PCE-based Path Computation Procedure.................5
  1.3. Traffic Engineering Aggregation and Abstraction.........5
  1.4. Traffic Engineered Label Switched Paths.................6
  1.5. Inter-area and Inter-AS Connectivity Discovery..........6
  1.6. Objective Functions.....................................6
2. Terminology...7
3. Issues and Considerations.......................................7
  3.1 Multi-homing...7
  3.2 Domain Confidentiality 8
  3.3 Destination Location...8
4. Domain Topologies...8
  4.1 Selecting Domain Paths.......................................9
  4.2 Multi-Homed Domains..9
  4.3 Domain Meshes..9
  4.4 Domain Diversity...9
  4.5 Synchronized Path Computations...............................9
  4.6 Domain Inclusion or Exclusion...............................10
5. Applicability of the PCE to Inter-area Traffic Engineering...10
  5.1. Inter-area Routing.....................................11
  5.1.1. Area Inclusion and Exclusion.........................11
  5.1.2. Strict Explicit Path and Loose Path..................11
  5.1.3. Inter-Area Diverse Path Computation..................12
  5.2. Control and Recording of Area Crossing.................12
6. Applicability of the PCE to Inter-AS Traffic Engineering.....12
  6.1. Inter-AS Routing.......................................13
  6.1.1. AS Inclusion and Exclusion...........................13
  6.1.2. Strict Explicit Path and Loose Path..................13

6.1.3. AS Inclusion and Exclusion............................13

6.2. Inter-AS Bandwidth Guarantees..........................13
6.3. Inter-AS Recovery.......................................14
6.4. Inter-AS PCE Peering Policies...........................14
7. Multi-Domain PCE Deployment...................................14
7.1 Traffic Engineering Database.............................14
7.2 Provisioning Techniques..................................14
7.3 Pre-Planning and Management-Based Solutions..............16
7.4 Per-Domain Computation...................................16
7.5 Cooperative PCEs...16
7.6 Hierarchical PCEs 16
8. Domain Confidentiality..17
8.1 Loose Hops...17
8.2 Confidential Path Segments and Path Keys.................17
9. Point-to-Multipoint...17
10. Optical Domains..18
10.1. PCE applied to the ASON Architecture...................18
11. Policy...19
12. TED Topology and Synchronization.............................19
12.1. Applicability of BGP-LS to PCE.........................20
13. Manageability Considerations.................................20
13.1 Control of Function and Policy..........................20
13.2 Information and Data Models.............................21
13.3 Liveness Detection and Monitoring.......................21
13.4 Verifying Correct Operation.............................21
13.5 Impact on Network Operation.............................21
14. Security Considerations......................................21
15. IANA Considerations..22
16. Acknowledgements...22
17. References...22
17.1. Normative References...................................22
17.2. Informative References.................................22
18. Author's Addresses...25

1. Introduction

Computing paths across large multi-domain environments may require special computational components and cooperation between entities in different domains capable of complex path computation. The Path Computation Element (PCE) [RFC4655] provides an architecture and a set of functional components to address this problem space.

A PCE may be used to compute end-to-end paths across multi-domain environments using a per-domain path computation technique [RFC5152]. The so called backward recursive path computation (BRPC) mechanism [RFC5441] defines a PCE-based path computation procedure to compute inter-domain constrained Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineered (TE) networks. However, both per-domain and BRPC techniques assume that the sequence of domains to be crossed from source to destination is known, either fixed by the network operator or obtained by other means.

In more advanced deployments (including multi-area and multi-AS environments) the sequence of domains may not be known in advance and the choice of domains in the end-to-end domain sequence might be critical to the determination of an optimal end-to-end path. In this case the use of the Hierarchical PCE [RFC6805] architecture and mechanisms may be used to discovery the intra-area path and select the optimal end-to-end domain sequence.

This document describes the processes and procedures available when using the PCE architecture, protocols and protocol extensions for computing inter-area and inter-AS MPLS and GMPLS Traffic TE paths.

1.1 Domains

A domain can be defined as a separate administrative, geographic, or switching environment within the network. A domain may be further defined as a zone of routing or computational ability. Under these definitions a domain might be categorized as an Antonymous System (AS) or an Interior Gateway Protocol (IGP) area (as per [RFC4726] and [RFC4655]).

For the purposes of this document, a domain is considered to be a collection of network elements within an area or AS that has a common sphere of address management or path computational responsibility. Wholly or partially overlapping domains are not within the scope of this document.

In the context of GMPLS, a particularly important example of a domain is the Automatically Switched Optical Network (ASON) subnetwork [G-8080]. In this case, computation of an end-to-end path requires the selection of nodes and links within a parent domain where some nodes may, in fact, be subnetworks. Furthermore, a domain might be an ASON routing area [G-7715]. A PCE may perform the path computation function of an ASON routing controller as described in [G-7715-2].

It is assumed that the PCE architecture should only be applied to small inter-domain topologies and not to solve route computation issues across large groups of domains, i.e. the entire Internet.

1.2 Path Computation

For the purpose of this document it is assumed that the path computation is the sole responsibility of the PCE as per the architecture defined in [RFC4655]. When a path is required the Path Computation Client (PCC) will send a request to the PCE. The PCE will apply the required constraints and compute a path and return a response to the PCC. In the context of this document it maybe necessary for the PCE to co-operate with other PCEs in adjacent domains (as per BRPC [RFC5441]) or cooperate with a Parent PCE (as per [RFC6805]).

It is entirely feasible that an operator could compute a path across multiple domains without the use of a PCE if the relevant domain information is available to the network planner or network management platform. The definition of what relevant information is required to perform this network planning operation and how that information is discovered and applied is outside the scope of this document.

1.2.1 PCE-based Path Computation Procedure

As highlighted, the PCE is an entity capable of computing an inter-domain TE path upon receiving a request from a PCC. There could be a single PCE per domain, or single PCE responsible for all domains. A PCE may or may not reside on the same node as the requesting PCC. A path may be computed by either a single PCE node or a set of distributed PCE nodes that collaborate during path computation.

[RFC4655] defines that a PCC should send a path computation request to a particular PCE, using [RFC5440] (PCC-to-PCE communication). This negates the need to broadcast a request to all the PCEs. Each PCC can maintain information about the computation capabilities of the PCEs it is aware of. The PCC-PCE capability awareness can configured using static configuration or by listening to the periodic advertisements generated by PCEs.

Once a path computation request is received, the PCC will send a request to the PCE. A PCE may compute the end-to-end path if it is aware of the topology and TE information required to compute the entire path. If the PCE is unable to compute the entire path, the PCE architecture provides co-operative PCE mechanisms for the resolution of path computation requests when an individual PCE does not have sufficient TE visibility.

A PCE may cooperate with other PCEs to determine intermediate loose hops. End-to-end path segments may be kept confidential through the application of path keys, to protect partial or full path information. A path key that is a token that replaces a path segment in an explicit route. The path key mechanism is described in [RFC5520]

1.3 Traffic Engineering Aggregation and Abstraction

Networks are often constructed from multiple areas or ASes that are interconnected via multiple interconnect points. To maintain network confidentiality and scalability TE properties of each area and AS are not generally advertized outside each specific area or AS.

TE aggregation or abstraction provide mechanism to hide information but may cause failed path setups or the selection of suboptimal

end-to-end paths [RFC4726]. The aggregation process may also have significant scaling issues for networks with many possible routes and multiple TE metrics. Flooding TE information breaks confidentiality and does not scale in the routing protocol.

The PCE architecture and associated mechanisms provide a solution to avoid the use of TE aggregation and abstraction.

1.4 Traffic Engineered Label Switched Paths

This document highlights the PCE techniques and mechanisms that exist for establishing TE packet and optical LSPs across multiple areas (inter-area TE LSP) and ASes (inter-AS TE LSP). In this context and within the remainder of this document, we consider all LSPs to be constraint-based and traffic engineered.

Three signaling options are defined for setting up an inter-area or inter-AS LSP [RFC4726]:

- Contiguous LSP
- Stitched LSP
- Nested LSP

All three signaling methods are applicable to the architectures and procedures discussed in this document.

1.5 Inter-area and Inter-AS Connectivity Discovery

When using a PCE-based approach for inter-area and inter-AS path computation, a PCE in one area or AS may need to learn information related to inter-AS capable PCEs located in other ASes. The PCE discovery mechanism defined in [RFC5088] and [RFC5089] allow the discovery of PCEs and disclosure of information related to inter-area and inter-AS capable PCEs across area and AS boundaries.

1.6 Objective Functions

An Objective Function (OF) [RFC5541], or set of OFs, specify the intentions of the path computation and so define the "optimality" in the context of that computation request.

An OF specifies the desired outcome of a computation. An OF does not describe or specify the algorithm to use, and an implementation may apply any algorithm or set of algorithms to achieve the result indicated by the OF. [RFC5541] provides the following OFs when computing inter-domain paths:

o Minimum Cost Path (MCP);
o Minimum Load Path (MLP);

o Maximum residual Bandwidth Path (MBP);
o Minimize aggregate Bandwidth Consumption (MBC);
o Minimize the Load of the most loaded Link (MLL);
o Minimize the Cumulative Cost of a set of paths (MCC).

OFs can be included in the PCE computation requests to satisfy the policies encoded or configured at the PCC, and a PCE may be subject to policy in determining whether it meets the OFs included in the computation request, or applies its own OFs.

During inter-domain path computation, the selection of a domain sequence, the computation of each (per-domain) path fragment, and the determination of the end-to-end path may each be subject to different OFs and policy.

2. Terminology

This document also uses the terminology defined in [RFC4655] and [RFC5440]. Additional terminology is defined below:

ABR: IGP Area Border Router, a router that is attached to more than one IGP area.

ASBR: Autonomous System Border Router, a router used to connect together ASes of a different or the same Service Provider via one or more inter-AS links.

CSPF: Constrained Shortest Path First.

Inter-area TE LSP: A TE LSP whose path transits through two or more IGP areas.

Inter-AS MPLS TE LSP: A TE LSP whose path transits through two or more ASes or sub-ASes (BGP confederations

SRLG: Shared Risk Link Group.

TED: Traffic Engineering Database, which contains the topology and resource information of the domain. The TED may be fed by Interior Gateway Protocol (IGP) extensions or potentially by other means.

3. Issues and Considerations

3.1 Multi-homing

Networks constructed from multi-areas or multi-AS environments may have multiple interconnect points (multi-homing). End-to-end path computations may need to use different interconnect points to avoid single point failures disrupting primary and backup services.

Domain and path diversity may also be required when computing end-to-end paths. Domain diversity should facilitate the selection of paths that share ingress and egress domains, but do not share transit domains. Therefore, there must be a method allowing the inclusion or exclusion of specific domains when computing end-to-end paths.

3.2 Domain Confidentiality

Where the end-to-end path crosses multiple domains, it may be possible that each domain (AS or area) are administered by separate Service Providers, it would break confidentiality rules for a PCE to supply a path segment to a PCE in another domain, thus disclosing AS-internal topology information.

If confidentiality is required between domains (ASes and areas) belonging to different Service Providers. Then cooperating PCEs cannot exchange path segments or else the receiving PCE PCC will be able to see the individual hops through another domain.

3.3 Destination Location

The PCC asking for an inter-domain path computation is typically aware of the identity of the destination node. Additionally, if the PCC is aware of the destination domain, it can supply this information as part of the path computation request. However, if the PCC does not know the egress domain this information must be determined by another method.

4. Domain Topologies

Constraint-based inter-domain path computation is a fundamental requirement for operating traffic engineered MPLS [RFC3209] and GMPLS [RFC3473] networks, in inter-area and inter-AS (multi-domain) environments. Path computation across multi-domain networks is complex and requires computational co-operational entities like the PCE.

4.1 Selecting Domain Paths

Where the sequence of domains is known a priori, various techniques can be employed to derive an optimal multi-domain path. If the domains are simply-connected, or if the preferred points of interconnection are also known, the Per-Domain Path Computation [RFC5152] technique can be used. Where there are multiple connections between domains and there is no preference for the choice of points of interconnection, BRPC [RFC5441] can be used to derive an optimal path.

When the sequence of domains is not known in advance, the optimum end-to-end path can be derived through the use of a hierarchical relationship between domains [RFC6805].

4.2 Multi-Homed Domains

Networks constructed from multi-areas or multi-AS environments may have multiple interconnect points (multi-homing). End-to-end path computations may need to use different interconnect points to avoid single point failures disrupting primary and backup services.

4.3 Domain Meshes

Very frequently network domains are composed by dozens or hundreds of network elements. These network elements are usually interconnected between them in a partial-mesh fashion, to provide survivability against dual failures, and to benefit from the traffic engineering capabilities from MPLS and GMPLS protocols. A typical node degree ranges from 3 to 10 (4-5 is quite common), being the node degree the number of neighbors per node.

Networks are sometimes divided into domains. Some reasons for it range from manageability to separation into vendor-specific domains. The size of the domain will be usually limited by control plane, but it can also be stated by arbitrary design constraints.

4.4 Domain Diversity

Whenever an specific connectivity service is required to have 1+1 protection feature, two completely disjoint paths must be established on an end to end fashion. In a multi-domain environment without, this can be accomplished either by selecting domain diversity, or by ensuring diverse connection within a domain. In order to compute the route diversity, it could be helpful to have SRLG information in the domains.

4.5 Synchronized Path Computations

In some scenarios, it would be beneficial for the operator to rely on the capability of the PCE to perform synchronized path computation.

Synchronized path computations, known as Synchronization VECtors (SVECs) are used for dependent path computations. SVECs are defined in [RFC5440] and [RFC6007] provides an overview for the use of the PCE SVEC list for synchronized path computations when computing dependent requests.

A non-comprehensive list of synchronized path computations include the following examples:

- o Route diversity: computation of two disjoint paths from a source to a destination (as drafted in the previous section).

- o Synchronous restoration: joint computation of a set of alternative paths for a set of affected LSPs as a consequence of a failure event. Note that in this case, the requests will potentially involve different source-destination pairs. In this scenario, the different path computation requests may arrive at different time stamps.

- o Batch provisioning: It is common that the operator sends a set of LSPs requests together, e.g in a daily of weekly basis, mainly in case of long lived LSPs. In order to optimize the resource usage, a synchronized path computation is needed.

- o Network optimization: After some time of operation, the distribution of the established LSP paths results in a non optimal use of resources. Also, inter-domain policies/agreements may have been changed. In such cases, a full (or partial) network planning action regarding the inter-domain connections will be triggered. This will involve the request of potentially a big amount of connections.

4.6 Domain Inclusion or Exclusion

A domain sequence is an ordered sequence of domains traversed to reach the destination domain, a domain sequence may be supplied during path computation to guide the PCEs or derived via use of Hierarchical PCE (H-PCE).

During multi-domain path computation, a PCC may request specific domains to be included or excluded in the domain sequence using the Include Route Object (IRO) [RFC5440] and Exclude Route Object (XRO) [RFC5521]. The use of Autonomous Number (AS) as an abstract node representing a domain is defined in [RFC3209], [DOMAIN-SEQ] specifies new sub-objects to include or exclude domains such as an IGP area or an Autonomous Systems.

5. Applicability of the PCE to Inter-area Traffic Engineering

As networks increase in size and complexity it may be required to introduce scaling methods to reduce the amount information flooded within the network and make the network more manageable. An IGP hierarchy is designed to improve IGP scalability by dividing the IGP domain into areas and limiting the flooding scope of topology information to within area boundaries. This restricts visibility of the area to routers in a single area. If a router needs to compute a route to destination located in another area a method is required to compute a path across area boundaries.

In order to support multiple vendors in a network, in cases where data and/or control plane technologies cannot interoperate, it is useful to divide the network in vendor domains. Each vendor domain is an IGP area, and the flooding scope of the topology (as well as any other relevant information) is limited to the area boundaries.

Per-domain path computation [RFC5152] exists to provide a method of inter-area path computation. The per-domain solution is based on loose hop routing with an Explicit Route Object (ERO) expansion on each Area Border Router (ABR). This allows an LSP to be established using a constrained path, however at least two issues exist:

- This method does not guarantee an optimal constrained path.

- The method may require several crankback signaling messages increasing signaling traffic and delaying the LSP setup.

The PCE-based architecture [RFC4655] is designed to solve inter-area path computation problems. The issue of limited topology visibility is resolved by introducing path computation entities that are able to cooperate in order to establish LSPs with source and destinations located in different areas.

5.1. Inter-area Routing

An inter-area TE-LSP is an LSP that transits through at least two IGP areas. In a multi-area network, topology visibility remains local to a given area, and a node in one area will not be able to compute an end-to-end path across multiple areas without the use of a PCE.

5.1.1. Area Inclusion and Exclusion

[RFC5152] provides the mechanisms to compute an inter-area path. It uses loose hop routing with an ERO expansion on each ABR. This allows the end-to-end path to be set up following a constrained path, but faces two major limitations:

- The method does not guarantee the use of an optimal constrained path.

- This may lead to several crankback signaling messages and hence delay the path setup.

[RFC5441] provides a more optimal method to specify inclusion or exclusion of an ABR. Using this method, an operator might decide if an area must be include or exclude from the inter-area path computation.

5.1.2. Strict Explicit Path and Loose Path

A strict explicit Path is defined as a set of strict hops, while a loose path is defined as a set of at least one loose hop and zero, one or more strict hops. It may be useful to indicate, during the path computation request, if a strict explicit path is required or not. An inter-area path may be strictly explicit or loose (e.g., a list of ABRs as loose hops).

A PCC request to a PCE does allow the indication of if a strict explicit path across specific areas is required or desired, or if the path request is loose.

5.1.3. Inter-Area Diverse Path Computation

It may be necessary (for protection or load-balancing) to compute a path that is diverse, from a previously computed path. There are various levels of diversity in the context of an inter-area network:

- Per-area diversity (intra-area path segments are link, node or SRLG disjoint.

- Inter-area diversity (end-to-end inter-area paths are link, node or SRLG disjoint).

Note that two paths may be disjoint in the backbone area but non-disjoint in peripheral areas. Also two paths may be node disjoint within areas but may share ABRs, in which case path segments within an area are node disjoint but end-to-end paths are not node-disjoint.

Both Per-Domain [RFC5152] and BRPC [RFC5441] mechanisms support the capability to compute diverse across multi-area topologies.

5.2. Control and Recording of Area Crossing

In some environments it be useful for the PCE to provide a PCC the set of areas crossed by the end-to-end path. Additionally the PCE can provide the path information and mark each segment so the PCC has visibility of which piece of the path lies within which area. Although by implementing Path-Key, the hop-by-hop (area topology) information is kept confidential.

6. Applicability of the PCE to Inter-AS Traffic Engineering

As discussed in section 4 (Applicability of the PCE to Inter-area Traffic Engineering) it is necessary to divide the network into smaller administrative domains, or ASes. If an LSR within an AS needs to compute a path across an AS boundary it must also use an inter-AS computation technique. [RFC5152] defines mechanisms for the computation of inter-domain TE LSPs using network elements along the signaling paths to compute per-domain constrained path segments.

The PCE was designed to be capable of computing MPLS and GMPLS paths across AS boundaries. This section outlines the features of a PCE-enabled solution for computing inter-AS paths.

6.1 Inter-AS Routing

6.1.1. AS Inclusion and Exclusion

[RFC5441] a method to specify inclusion or exclusion of an ASBR. Using this method, an operator might decide if an AS must be include or exclude from the inter-AS path computation.

6.1.2. Strict Explicit Path and Loose Path

During path computation, the PCE architecture and BRPC algorithm allow operators to specify if the resultant LSP must follow a strict or a loose path. By explicitly specify the path, the operator request a strict explicit path which must pass through one or many LSR. If this behaviour is well define and appropriate for inter-area, it implies some topology discovery for inter-AS. So, this feature when the operator owns several ASes (and so, knows the topology of its ASes) or restricts to the well-known ASBR to avoid topology discovery between operators. The loose path, even if it does not allow granular specification of the path, protects topology disclosure as it not obligatory for the operator to disclose information about its networks.

6.1.3. AS Inclusion and Exclusion

Like explicit and loose path, [RFC5441] allows to specify inclusion or exclusion of respectively an AS or an ASBR. Using this method, an operator might decide if an AS must be include or exclude from the inter-AS path computation. Exclusion and/or inclusion could also be specified at any step in the LSP path computation process by a PCE (within the BRPC algorithm) but the best practice would be to specify them at the edge. In opposition to the strict and loose path, AS inclusion or exclusion doesn't impose topology disclosure as ASes are public entity as well as their interconnection.

6.2 Inter-AS Bandwidth Guarantees

Many operators with multi-AS domains will have deployed MPLS-TE DiffServ either across their entire network or at the domain edges on CE-PE links. In situations where strict QOS bounds are required, admission control inside the network may also be required.

When the propagation delay can be bounded, the performance targets, such as maximum one-way transit delay may be guaranteed by providing bandwidth guarantees along the DiffServ-enabled path.

One typical example of this requirement is to provide bandwidth guarantees over an end-to-end path for VoIP traffic classified as EF (Expedited Forwarding) class in a DiffServ-enabled network. In the case where the EF path is extended across multiple ASes, inter-AS bandwidth guarantee would be required.

Another case for inter-AS bandwidth guarantee is the requirement for guaranteeing a certain amount of transit bandwidth across one or multiple ASes.

6.3 Inter-AS Recovery

During path computation, a PCC request may contain backup LSP requirements in order to setup in the same time the primary and backup LSPs. It is also possible to request a backup LSP for a group of primary LSPs. [RFC4090] adds fast re-route protection to LSP. So, the PCE could be used to trigger computation of backup tunnels in order to protect Inter-AS connectivity. Inter-AS recovery requirements needs not only PCE protection and redundancy but also LSP tunnels protection through FRR mechanisms. Inter-AS PCE computation must support the FRR mechanisms and the patch computation for backup tunnels for protection and fast recovery.

6.4 Inter-AS PCE Peering Policies

Like BGP peering policies, inter-AS PCE peering policies is a requirement for operator. In inter-AS BRPC process, PCE must cooperate in order to compute the end-to-end LSP. So, the AS path must not only follow technical constraints e.g. bandwidth availability, but also policies define by the operator.

Typically PCE interconnections at an AS level must follow contract obligations, also known as peering agreements. The PCE peering policies are the result of the contract negotiation and govern the relation between the different PCE.

7. Multi-domain PCE Deployment Options

The PCE provides the architecture and mechanisms to compute inter-area and inter-AS LSPs. The objective of this document is not to reprint the techniques and mechanisms available, but to highlight their existence and reference the relevant documents that introduce and describe the techniques and mechanisms necessary for computing inter-area and inter-AS LSP based services.

An area or AS may contain multiple PCEs:

- The path computation load may be balanced among a set of PCEs to improve scalability.

- For the purpose of redundancy, primary and backup PCEs may be used.

- PCEs may have distinct path computation capabilities (P2P or P2MP).

Discovery of PCEs and capabilities per area or AS is defined in [RFC5088] and [RFC5089].

Each PCE per domain can be deployed in a centralized or distributed architecture, the latter model having local visibility and collaborating in a distributed fashion to compute a path across the domain. Each PCE may collect topology and TE information from the same sources as the LSR, such as the IGP TED.

When the PCC sends a path computation request to the PCE, the PCE will compute the path across a domain based on the required constraints. The PCE will generate the full set of strict hops from source to destination. This information, encoded as an ERO, is then sent back to the PCC that requested the path. In the event that a path request from a PCC contains source and destination nodes that are located in different domains the PCE is required to co-operate between multiple PCEs, each responsible for its own domain.

Techniques for inter-domain path computation are described in [RFC5152] and [RFC5441], both techniques assume that the sequence of domains to be crossed from source to destination is well known. In the event that the sequence of domains is not well known, [RFC6805] might be used. The sequence could also be retrieve locally from information previously stored in the PCE database (preferably in the TED) by OSS management or other protocols.

7.1 Traffic Engineering Database

TEDs are automatically populated by the IGP-TE like IS-IS-TE or OSPF-TE. However, no information related to AS path are provided by such IGP-TE. It could be helpful for BRPC algorithm as AS path helper, to populate a TED with suitable information regarding inter-AS connectivity. Such information could be obtain from various sources, such as BGP protocol, peering policies, OSS of the operator or from neighbor PCE. In any case, no topology disclosure must be impose in order to provide such information.

In particular, for both inter-area and inter-AS, the TED must be populated with all boundary node information suitable to establish PCEP protocol with the next PCE in the path.

7.2 Provisioning Techniques

As PCE algorithms rely on information contained in the TED, it is possible to populate TED information by means of provisioning. In

this case, the operator must regularly update and store all suitable information in the TED in order for the PCE to correctly compute LSP. Such information range from policies (e.g. avoid this LSR, or use this ASBR for a specific IP prefix) up to topology information (e.g. AS X is reachable trough a 100 Mbit/s link on this ASBR and 30 Mbit/s are reserved for EF traffic). Operators may choose the type and amount of information they can use to manage their traffic engineered network.

However, some LSPs might be provisioned to link ASes or areas. In this case, these LSP must be announced by the IGP-TE in order to automatically populate the TED.

7.3 Pre-Planning and Management-Based Solutions

Offline path computation is performed ahead of time, before the LSP setup is requested. That means that it is requested by, or performed as part of, a management application. This model can be seen in Section 5.5 of [RFC4655].

The offline model is particularly appropriate to long-lived LSPs (such as those present in a transport network) or for planned responses to network failures. In these scenarios, more planning is normally a feature of LSP provisioning.

This model may also be used where the network operator wishes to retain full manual control of the placement of LSPs, using the PCE only as a computation tool to assist the operator, not as part of an automated network.

The management based solutions could also be used in conjunction with the BRPC algorithm. Operator just computes the AS-Path as parameter for the inter-AS path computation request and let each PCE along the AS path compute the LSP part on its own domain.

7.4 Per-Domain Computation

[RFC5152] defines the mechanism to compute per-domain path and must be used in that condition. Otherwise, BRPC [RFC5441] will be used.

7.5 Cooperative PCEs

When PCE cooperate to compute an inter-area or inter-AS LSP, both [RFC5152] and [RFC5441] could be used.

7.6 Hierarchical PCEs

The [RFC6805] draft defines how a hierarchy of PCEs may be used. An operator must define a parent PCE and each child PCE. A parent PCE

can be announced in the other areas or ASes in order for the parent PCE to contact remote child PCEs. Reciprocally, child PCEs are announced in remote areas or ASes in order to be contacted by a remote parent PCE. Parent and each child PCE could also be provisioned in the TED if they are not announced.

8. Domain Confidentiality

Confidentiality typically applies to inter-provider (inter-AS) PCE communication. Where the TE LSP crosses multiple domains (ASes or areas), the path may be computed by multiple PCEs that cooperate together. With each local PCE responsible for computing a segment of the path. However, in some cases (e.g., when ASes are administered by separate Service Providers), it would break confidentiality rules for a PCE to supply a path segment to a PCE in another domain, thus disclosing AS-internal or area topology information.

8.1 Loose Hops

A method for preserving the confidentiality of the path segment is for the PCE to return a path containing a loose hop in place of the segment that must be kept confidential. The concept of loose and strict hops for the route of a TE LSP is described in [RFC3209].

[RFC5440] supports the use of paths with loose hops, and it is a local policy decision at a PCE whether it returns a full explicit path with strict hops or uses loose hops. A path computation request may request an explicit path with strict hops, or may allow loose hops as detailed in [RFC5440].

8.2 Confidential Path Segments and Path Keys

[RFC5520] defines the concept and mechanism of Path-Key. A Path-Key is a token that replaces the path segment information in an explicit route. The Path-Key allows the explicit route information to be encoded and in the PCEP ([RFC5440]) messages exchanged between the PCE and PCC.

This Path-Key technique allows explicit route information to used for end-to-end path computation, without disclosing internal topology information between domains.

9. Point-to-Multipoint

For the Point-to-Multipoint application scenarios for MPLS-TE LSP, the complexity of domain sequences, domain policies, choice and number of domain interconnects is magnified comparing to P2P path

computations. Also as the size of the network grows, the number of leaves and branches increase and it in turn puts the scalability of the path computation and optimization into a bigger issue. A solution for the point-to-multipoint path computations may be achieved using the PCEP protocol extension for P2MP [RFC6006] and using the PCEP P2MP procedures defined in [PCEP-P2MP-INTER-DOMAIN].

10. Optical Domains

The International Telecommunications Union (ITU) defines the ASON architecture in [G-8080]. [G-7715] defines the routing architecture for ASON and introduces a hierarchical architecture. In this architecture, the Routing Areas (RAs) have a hierarchical relationship between different routing levels, which means a parent (or higher level) RA can contain multiple child RAs. The interconnectivity of the lower RAs is visible to the higher level RA.

10.1. PCE applied to the ASON Architecture

In the ASON framework, a path computation request is termed a Route Query. This query is executed before signaling is used to establish an LSP termed a Switched Connection (SC) or a Soft Permanent Connection (SPC). [G-7715-2] defines the requirements and architecture for the functions performed by Routing Controllers (RC) during the operation of remote route queries - an RC is synonymous with a PCE.

In the ASON routing environment, a RC responsible for an RA may communicate with its neighbor RC to request the computation of an end-to-end path across several RAs. The path computation components and sequences are defined as follows:

- o Remote route query. An operation where a routing controller communicates with another routing controller, which does not have the same set of layer resources, in order to compute a routing path in a collaborative manner.

- o Route query requester. The connection controller or RC that sends a route query message to a routing controller requesting for one or more routing path that satisfies a set of routing constraints.

- o Route query responder. An RC that performs path computation upon reception of a route query message from a routing controller or connection controller, sending a response back at the end of computation.

When computing an end-to-end connection, the route may be computed by a single RC or multiple RCs in a collaborative manner and the two

scenarios can be considered a centralized remote route query model and distributed remote route query model. RCs in an ASON environment can also use the hierarchical PCE [RFC6805] model to fully match the ASON hierarchical routing model.

11. Policy

Policy is important in the deployment of new services and the operation of the network. [RFC5394] provides a framework for PCE-based policy-enabled path computation. This framework is based on the Policy Core Information Model (PCIM) as defined in [RFC3060] and further extended by [RFC3460].

When using a PCE to compute inter-domain paths, policy may be invoked by specifying:

- Each PCC must select which computations will be delegated to a PCE;

- Each PCC must select which PCEs it will use;

- Each PCE must determine which PCCs are allowed to use its services and for what computations;

- The PCE must determine how to collect the information in its TED, who to trust for that information, and how to refresh/update the information;

- Each PCE must determine which objective functions and which algorithms to apply.

Finally, due to the nature of inter-domain (and particularly using H-PCE based) path computations, deployment of policy should also consider the need to be sensitive to commercial and reliability information about domains and the interactions of services crossing domains.

12. TED Topology and Synchronization

The PCE operates on a view of the network topology as presented by a Traffic Engineering Database. As discussed in [RFC4655] the TED used by a PCE may be learnt by the relevant IGP extensions.

Thus, the PCE may operate its TED is by participating in the IGP running in the network. In an MPLS-TE network, this would require OSPF-TE [RFC3630] or ISIS-TE [RFC5305]. In a GMPLS network it would utilize the GMPLS extensions to OSPF and IS-IS defined in [RFC4203] and [RFC5307].

An alternative method to provide network topology and resource information is offered by [BGP-LS], which is described in the following section.

12.1 Applicability of BGP-LS to PCE

The concept of exchange of TE information between Autonomous Systems (ASes) is discussed in [BGP-LS]. The information exchanged in this way could be the full TE information from the AS, an aggregation of that information, or a representation of the potential connectivity across the AS. Furthermore, that information could be updated frequently (for example, for every new LSP that is set up across the AS) or only at threshold-crossing events.

There are a number of discussion points associated with the use of [BGP-LS] concerning the volume of information, the rate of churn of information, the confidentiality of information, the accuracy of aggregated or potential-connectivity information, and the processing required to generate aggregated information. The PCE architecture and the architecture enabled by [BGP-LS] make different assumptions about the operational objectives of the networks, and this document does not attempt to make one of the approaches "right" and the other "wrong". Instead, this work assumes that a decision has been made to utilize the PCE architecture.

Indeed, [BGP-LS] may have some uses within the PCE model. For example, [BGP-LS] could be used as a "northbound" TE advertisement such that a PCE does not need to listen to an IGP in its domain, but has its TED populated by messages received (for example) from a Route Reflector. Furthermore, the inter-domain connectivity and connectivity capabilities that is required optional information for a parent PCE could be obtained as a filtered subset of the information available in [BGP-LS].

13. Manageability Considerations

General PCE management considerations are discussed in [RFC4655]. In the case of multi-domains within a single service provider network, the management responsibility for each PCE would most likely be handled by the same service provider. In the case of multiple ASes within different service provider networks, it will likely be necessary for each PCE to be configured and managed separately by each participating service provider, with policy being implemented based on an a previously agreed set of principles.

13.1 Control of Function and Policy

As per PCEP [RFC5440] implementation allow the user to configure a number of PCEP session parameters. These are detailed in section 8.1 of [RFC5440] and will not be repeated here.

13.2 Information and Data Models

A PCEP MIB module is defined in [PCEP-MIB] that describes managed objects for modeling of PCEP communication including:

- o PCEP client configuration and status,
- o PCEP peer configuration and information,
- o PCEP session configuration and information,
- o Notifications to indicate PCEP session changes.

13.3 Liveness Detection and Monitoring

PCEP includes a keepalive mechanism to check the liveliness of a PCEP peer and a notification procedure allowing a PCE to advertise its overloaded state to a PCC. In a multi-domain environment [RFC5886] provides the procedures necessary to monitor the liveliness and performances of a given PCE chain.

13.4 Verifying Correct Operation

In order to verify the correct operation of PCEP, [RFC5440] specifies the monitoring of key parameters. These parameters are detailed in section 8.4 of [RFC5440] and will not be repeated here.

13.5 Impact on Network Operation

[RFC5440] states that in order to avoid any unacceptable impact on network operations, a PCEP implementation should allow a limit to be placed on the number of sessions that can be set up on a PCEP speaker, it may also be practical to place a limit on the rate of messages sent by a PCC and received my the PCE.

14. Security Considerations

PCEP security is defined [RFC5440]. Any multi-domain operation necessarily involves the exchange of information across domain boundaries. This does represent a significant security and confidentiality risk. PCEP allows individual PCEs to maintain confidentiality of their domain path information using path-keys [RFC5520].

For further considerations of the security issues related to inter-domain path computation, see [RFC5376].

15. IANA Considerations

This document makes no requests for IANA action.

16. Acknowledgements

The author would like to thank Adrian Farrel for his review, and Meral Shirazipour and Francisco Javier Jimenex Chico for their comments.

17. References

17.1. Normative References

17.2. Informative References

[RFC3060] Moore, B., Ellesson, E., Strassner, J., and A. Westerinen, "Policy Core Information Model -- Version 1 Specification", RFC 3060, February 2001.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3460] Moore, B., Ed., "Policy Core Information Model (PCIM) Extensions", RFC 3460, January 2003.

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

[RFC4203] Kompella, K., Ed., and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation

Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC4726] Farrel, A., Vasseur, J., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.

[RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

[RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.

[RFC5152] Vasseur, JP., Ayyangar, A., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

[RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

[RFC5376] Bitar, N., et al., "Inter-AS Requirements for the Path Computation Element Communication Protocol (PCECP)", RFC 5376, November 2008.

[RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, December 2008.

[RFC5440] Ayyangar, A., Farrel, A., Oki, E., Atlas, A., Dolganow, A., Ikejiri, Y., Kumaki, K., Vasseur, J., and J. Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

[RFC5441] Vasseur, J.P., Ed., "A Backward Recursive PCE-based Computation (BRPC) procedure to compute shortest inter-domain Traffic Engineering Label Switched Paths", RFC5441, April 2009.

[RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.

[RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the

Path Computation Element Communication Protocol (PCEP)
for Route Exclusions", RFC 5521, April 2009.

[RFC5541] Le Roux, J., Vasseur, J., Lee, Y., "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC5541, December 2008.

[RFC5886] Vasseur, JP., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path ComputationElement (PCE)-Based Architecture", RFC 5886, June 2010.

[RFC6006] Takeda, T., Chaitou M., Le Roux, J.L., Ali Z., Zhao, Q., King, D., "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC6006, September 2010.

[RFC6007] Nishioka, I., King, D., "Use of the Synchronization VECtor (SVEC) List for Synchronized Dependent Path Computations", RFC6007, September 2010.

[G-8080] ITU-T Recommendation G.8080/Y.1304, Architecture for the automatically switched optical network (ASON).

[G-7715] ITU-T Recommendation G.7715 (2002), Architecture and Requirements for the Automatically Switched Optical Network (ASON).

[G-7715-2] ITU-T Recommendation G.7715.2 (2007), ASON routing architecture and requirements for remote route query.

[RFC6805] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS & GMPLS", RFC6805, July 2010.

[PCEP-P2MP-INTER-DOMAIN] Zhao, Q., Dhody, D., Ali Z., King, D., Casellas, R., "PCE-based Computation Procedure To Compute Shortest Constrained P2MP Inter-domain Traffic Engineering Label Switched Paths", work in progress.

[BGP-LS] Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE Information using BGP", work in progress.

[PCEP-MIB] Stephan, E., Koushik, K., Zhao, Q., King, D., "PCE Communication Protocol (PCEP) Management Information Base", work in progress.

[DOMAIN-SEQ] Dhody, D., Palle, U., and R. Casellas, "Standard Representation Of Domain Sequence", work in progress.

18. Author's Addresses

Daniel King
Old Dog Consulting
UK

EMail: daniel@olddog.co.uk

Julien Meuric
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex

EMail: julien.meuric@orange-ftgroup.com

Olivier Dugeon
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex

EMail: olivier.dugeon@orange-ftgroup.com

Quintin Zhao
Huawei Technology
125 Nagog Technology Park
Acton, MA 01719
US

EMail: qzhao@huawei.com

Oscar Gonzalez de Dios
Telefonica I+D
Emilio Vargas 6, Madrid
Spain

EMail: ogondio@tid.es