

PCEP Extension for Stateful Inter-Domain Tunnels
draft-dugeon-pce-stateful-interdomain-00

Abstract

The Path Computation Element (PCE) working group (WG) has produced a set of RFCs to standardize the behavior of the Path Computation Element as a tool to help MPLS-TE, GMPLS LSP tunnels and Segment Routing paths placement. This also include the ability to compute inter-domain LSPs or Segment Routing path following a distributed or hierarchical approach. In complement to the original stateless mode, a stateful mode has been added. In particular, the new PCInitiate message allows a PCE to directly ask a PCC to setup an MPLS-TE, GMPLS LSP tunnel or a Segment Routing path. However, once computed, the inter-domain LSPs or Segment Routing path are hard to setup in the underlying network. Especially, in operational network, RSVP-TE signaling is not enable between AS border routers. But, such RSVP-TE signaling is mandatory to setup contiguous LSP tunnels or to stitch or nest independent LSP tunnels to form the end-to-end inter-domain paths. This document proposes to combine a Backward Recursive or Hierarchical method with PCInitiate message to setup independent paths per domain, and combine these different paths together in order to operated them as end-to-end inter-domain paths without the need of signaling session between AS border routers. A new Stitching Label is defined and new LSP-TYPE code points are considered for that purpose.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 30, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Problem Statement	3
1.1.	General assumptions	4
1.2.	Terminology	5
2.	Stitching Label	7
2.1.	Definition	7
2.2.	Inter-domain LSP-TYPE	7
3.	Backward Recursive PCInitiate procedure	8
3.1.	Mode of operation	8
3.2.	Example	10
3.3.	Inter-domain LSP setup procedure completion failure . . .	12
4.	Hierarchical PCInitiate procedure	13
4.1.	Mode of operation	13
4.2.	Inter-domain LSP setup procedure completion failure . . .	15
5.	Inter-domain LSP Management	16
5.1.	Identification of inter-domain tunnels	16
5.2.	Inter-domain LSP management	16
5.3.	Modification of inter-domain LSP	17
5.4.	Removal of inter-domain LSP	17
6.	Applicability	18
7.	IANA Considerations	19
7.1.	LSP-TYPE values	19
7.2.	PCEP-Error Object	19
8.	Security Considerations	19
9.	Acknowledgements	19

10.	Disclaimer	20
11.	References	20
11.1.	Normative References	20
11.2.	Informative References	21
	Authors' Addresses	22

[1.](#) Problem Statement

Looking to the different RFCs that describe the PCE architecture and in particular PCE based architecture [[RFC4655](#)], PCE protocol [[RFC5440](#)], BRPC [[RFC5441](#)] and H-PCE [[RFC6805](#)], the Path Computation Element (PCE) is able to compute inter-domain paths in complement to intra-domain computation. Such inter-domain paths could then serve as the Explicit Route Object input for the RSVP-TE signaling to setup the LSPs tunnel within the underlying network. Three kinds of inter-domain paths could be established:

- o Contiguous LSPs ([[RFC3209](#)] and [[RFC3473](#)]): The RSVP-TE signaling crosses the boundary between two domains, e.g. between two AS Border Routers (ASBRs) as if they were two routers of the same domain. This kind of tunnel is not recommended mostly for security and scalability purpose. In addition, the initiating domain imposes huge constraints on subsequent domains, because they undergo the tunnel request without being able to control it.
- o Stitching LSPs ([[RFC5150](#)]): Each domain establishes in its own network the corresponding part of the inter-domain path independently. Then, a second end-to-end RSVP-TE Path message is sent by the initiating domain to stitch the different tunnel parts to form the inter-domain path. In fact, this second RSVP-TE Path message is used by border nodes to exchange the label that must be used by the previous domain to send the traffic in order that the MPLS packets follow the next LSP tunnel in the following domain. These labels are conveyed in the RSVP-TE Resv message.
- o Nesting LSPs ([[RFC4206](#)]): This is similar to the stitching mode but, this time, with the possibility to setup tunnel hierarchy. For example, an LSP tunnel between two edge domains crossing a transit domain could be carried over a tunnel of a higher level in the transit domain. Again, a second end-to-end RSVP-TE Path message is sent from the source to the destination. Labels that must be used by the ASBRs of transit domains to identify flows to be nested are carried by the RSVP-TE Resv message.

In all case, RSVP-TE signaling must be exchanged between the different domains. However, from an operational point of view, looking to different networks under the responsibility of different administrative entities, only BGP sessions are setup and configured

between ASBRs. Technologically speaking, this is possible and many RFCs describe how to use RSVP-TE for inter-domain. But, due to security, scalability, management and contract constraints, RSVP-TE is not exposed at the network boundary. To circumvent some of the security issues, RSVP-TE can be carried inside an IPsec tunnel between ASBRs, but, this does not eliminate the scalability aspect nor the constraints imposed by setting up inter-domain paths.

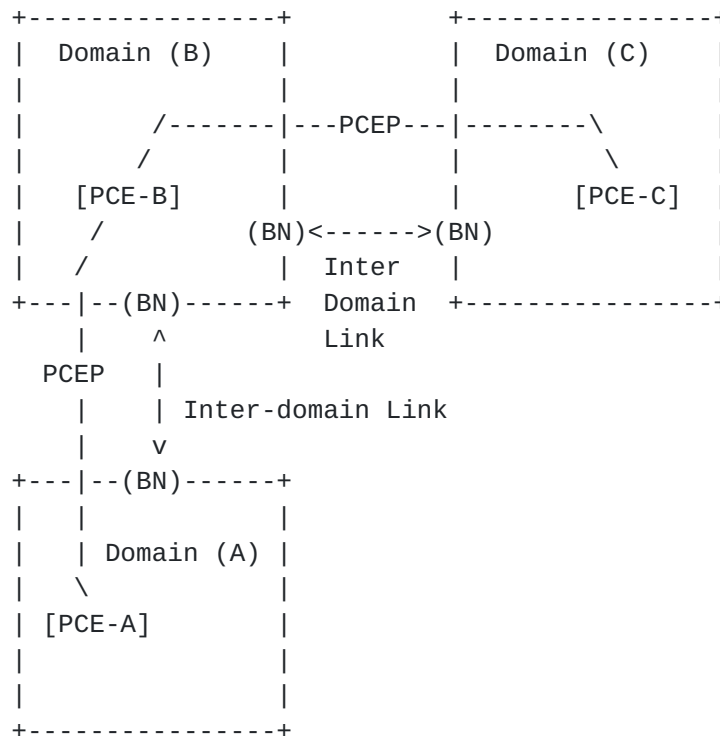
The purpose of this memo is to take the benefit of PCE Stateful [[RFC8231](#)] and PCE Initiated [[I-D.ietf-pce-pce-initiated-lsp](#)] modes to stitch or nest inter-domain paths directly using PCEP between domains' PCEs instead of using RSVP-TE signaling at the inter-domain, while keeping each operator free to independently setup their respective part of the inter-domain paths. PCInitiated message is used in a Backward Recursive way like the PCReq message in BRPC [[RFC5441](#)], to recursively setup the end-to-end tunnel. PCRep message is used to automatically stitch or nest the different local LSP tunnels. And, PCRep in conjunction of PCUpd messages are used to maintain, modify and remove inter-domain paths. This method is also applicable to Segment Routing to build inter-domain segment paths.

1.1. General assumptions

In the rest of this document, we used the same references as per BRPC [[RFC5441](#)] and make the following set of assumptions (see figure below):

- o Domain refers to an IGP area or an Autonomous System (AS).
- o Inter-domain path is used to refer to a path that cross two or more different domains as defined previously,
- o At least, one PCE is deployed in each domain. These PCE are all stateful active capable and could request to enforce LSP tunnels in their respective domain by means of PCInitiate messages.
- o LSRs, including border nodes, are PCC enable and support stateful active mode. PCEP sessions is established between these routers and their domains' PCE.
- o Each PCE establishes a PCEP session with its respective neighbor domain's PCE. The way a PCE discover its neighboring PCE is out of scope of this draft. These information could be fulfill administratively or automatically discovered through, for example per draft 'BGP Extensions for Path Computation Element (PCE) Discovery' [[I-D.dong-pce-discovery-proto-bgp](#)],

- o PCEs are able to compute and end-to-end path as per BRPC procedure [[RFC5441](#)].
- o Tunnels refer to LSPs setup by mean of RSVP-TE or Segment Path in a Segment Routing network.



Example of the representation of 3 domains with 3 PCEs

1.2. Terminology

ABR: Area Border Routers. Routers used to connect two IGP areas (areas in OSPF or levels in IS-IS).

ASBR: Autonomous System Border Router. Router used to connect together ASes of the same or different service providers via one or more inter-AS links.

AS: Autonomous System

Border Node (BN): a boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

Domains: Autonomous System (AS) or IGP Area. An Autonomous System is composed by one or more IGP area.

BNe(i): Entry BN of domain(i) connecting domain(i-1) to domain(i) along a determined sequence of domains. Multiple entry BN(i) could be used to connect domain(i-1) to domain(i).

BNo(i): Output BN of domain(i) connecting domain(i) to domain(i+1) along a determined sequence of domains. Multiple exit BN(i) could be used to connect domain(i) to domain(i+1).

Inter-domain path: A path that crosses two or more domains through a pair of Border Node (BNo, BNe).

Local LSP tunnel: A LSP tunnel that do not cross a domain. It is setup between entry BNe to output BNo, any source to output BNo or entry BNe to any destination of the same domain. This LSP could be enforce by means of RSVP-TE signaling or Segment Routing labels stack.

Local LSP tunnel(i): A local LSP tunnel of domain(i)

IGP-TE: Interior Gateway Protocol with Traffic Engineering support. Both OSPF-TE and IS-IS-TE are identified in this category.

Stitching Label (SL): A dedicated label that is used to stitch two RSVP-TE tunnels or two Segment Routing paths.

SL(i): A Stitching Label that link domain(i) to domain(i-1).

LK(i): A Link that connect BNo(i) to BNe(i+1). Note that BNo(i) could be connected to BNe(i+1) by more than one link. LK(i) serves to identify which of the multiple links will be used for the inter-domain LSP setup.

PLSP-ID(i): A PLSP-ID that identify the local tunnel part of an inter-domain tunnel in the domain(i).

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCE(i) is a PCE with the scope of domain(i).

2. Stitching Label

This section introduce the concept of Stitching Label that allows stitching and nesting of local LSP tunnels in order to form inter-domain path that cross several different domains.

2.1. Definition

The operation of stitch or nest a local LSP tunnel(i) to a local LSP tunnel(i+1) in order to form an inter-domain path simply consist in defining the label that the output BNo(i) will use to send its traffic to the entry BNe(i+1). Indeed, the entry BNe(i+1) needs to identify the incoming traffic i.e. IP packets, in order to know if this traffic must follow the local LSP tunnel(i+1) or not. Forwarding Equivalent Class (FEC) could be used for that purpose. But, when stitching or nesting tunnels, the FEC is reduce to the incoming label that the entry BNe(i+1) as chosen for the local LSP tunnel(i+1).

In this memo, we introduce the named of 'Stitching Label (SL)' to designate this label. Such label is usually exchange between output BNo(i) and entry BNe(i+1) with the RSVP-TE signaling. But, as we want to avoid to use RSVP-TE signaling due to operational constraints, this Stitching Label will be convey by PCEP protocol. In fact, the Explicit Route Object (ERO) and the Record Route Object (RRO) are defined in order to transport this Stitching Label in the RSVP-TE signaling. As PCEP protocol used RSVP-TE Objects, and in particular the ERO and RRO, it is able to convey the Stitching Label without any modification of the PCEP protocol nor the PCE or RSVP-TE Objects.

As per [RFC4003](#) [[RFC4003](#)], the Stitching Label will be convey as a companion of an IP address. In our case, this is one of the IP address of the link LK(i) which connects BNo(i) to BNe(i+1) and carries the traffic from the domain(i) to domain(i+1). It is left to implementation to select which of the two IP address of the link LK(i) is used.

2.2. Inter-domain LSP-TYPE

However, even if PCEP could convey the Stitching Label, a PCC is not aware that a PCE requests or provides such label. For that purpose, this memo propose to use the LSP-TYPE as defined in draft lsp setup type [[I-D.ietf-pce-lsp-setup-type](#)] with new values (See IANA section of this memo) defined as follow:

- o TBD1: Inter-Domain Traffic engineering end-to-end path is setup using Backward Recursive or Hierarchical method. This new LSP-

TYPE value MUST be set in a PCInitiate messages sends by a PCE(i) to its neighbor PCE(i+1) in the Backward Recursive method or by the Parent PCE to the Child PCE(i) to initiate a new inter-domain path. In turn, neighbor PCE(i+1) or Child PCE(i) MUST return a Stitching Label SL with the IP address of the associated link in the RRO of the PCRpt message to PCE(i) or Parent PCE.

- o TBD2: Inter-Domain Traffic engineering local path is setup using RSVP-TE. This new LSP-TYPE value MUST be set in the PCInitiate message sends by a PCE(i) requesting to a PCC of domain(i) to initiate a new local LSP tunnel(i) which is part of an inter-domain path. This LSP-TYPE value MUST be used by the PCE(i) only after receiving a PCInitiate message with an LSP-TYPE equal to TBD1 from a neighbor PCE(i+1) in the Backward Recursive method or Parent PCE in the Hierarchical method. In turn, the PCC of domain(i) MUST return a Stitching Label SL with the IP address of associated link in the RRO of the PCRpt message.
- o TBD3: Inter-Domain Traffic engineering local path is setup using Segment Routing. This new LSP-TYPE value MUST be set in the PCInitiate message sends by a PCE(i) requesting to a PCC of domain(i) to initiate a new Segment Routing path which is part of and inter-domain Segment Routing path. This LSP-TYPE value MUST be used by the PCE(i) only after receiving a PCInitiate message with an LSP-TYPE equal to TBD1 from a neighbor PCE(i+1). In turn, the PCC MUST return a Stitching Label SL with the IP address of the associated link in the RRO of the PCRpt message.

3. Backward Recursive PCInitiate procedure

This section describes how to setup inter-domain paths than cross several different domains by using a Backward Recursive method which is compatible to inter-domain path computation by means of the BRPC procedure as describe in [RFC5441](#) [[RFC5441](#)].

3.1. Mode of operation

This section describes how PCInitiate and PCRpt messages are combined between PCE in order to setup inter-domain paths between a source domain(1) to a destination domain(n). S and D are respectively the source and destination of the inter-domain path. Domain(1) and domain(n) are different and connected through 0 or more intermediate domains denoted domain(i) with $i = (2, n-1)$. Domains are directly connected when $n = 2$.

First, the PCE(S) run standard BRPC algorithm as per [RFC5441](#) [[RFC5441](#)] with its neighbor PCEs in order to compute the inter-domain path from S to D, where S and D are respectively a node in the

domain(1) and domain(n). Path Key confidentiality as per [RFC5520](#) [RFC5520] MAY be used to obfuscate the detailed ERO of the different domains(i). The resulting ERO is of the form {S, PKS(1), BNo(1), ..., BNe(i), PKS(i), BNo(i), ..., BNe(n), PKS(n), D}. As subsequent domains are not aware about the final computed ERO in case of multiple VSPT, the final computed ERO MUST be send in the PCInitiate message to indicate to the subsequent PCEs which solution has been finally chosen.

The complete procedure follow the different steps described below:

Steps 1: Initialization

Once ERO(S, D) computed, PCE(1) sends a PCInitiate message to PCE(2) containing and ERO equal to {S, PKS(1), BNo(1), ..., BNe(i), PKS(i), BNo(i), ..., BNe(n), PKS(n), D}, LSP-TYPE = TBD1 and End-Points Object = (S, D). The ERO corresponds to the one PCE(1) as received from PCE(2) during the BRPC process. In case of multiple EROs, i.e. VSPT > 1, PCE(1) has chosen one of them and used the selected one for the PCInitiate message. PKS(i) could be replaced by the full ERO description if Path Key is not used by PCE(i).

When PCE(i) receives a PCInitiate message from domain(i-1) with LSP-TYPE = TBD1 and ERO = {BNe(i), PKS(i), BNo(i), ..., BNe(n), PKS(n), D}, it forwards the PCInitiate message to PCE(i+1) once remove its {BNe(i), PKS(i), BNo(i)} part from the ERO. All PCE(i) propagate the PCInitiate message to PCE(i+1) up to PCE(n) i.e. to the domain(n).

Steps 2: Actions taken at the destination domain(n) by PCE(n)

When PCInitiate message propagation reach the destination domain(n), PCE(n) retrieves the ERO from the PKS(n) if necessary and sends to BNe(n) a PCInitiate message with the ERO(n) = {BNe(n), ..., D}, LSP-TYPE= TBD2 and End-Points Object = {BN(n), D} in order to inform the PCC BNe(n) that this local LSP tunnel(n) is part of an inter-domain path. When the PCC BNe(n) received the PCInitiate message from its PCE(n), it setup the local LSP tunnels from entry BNe(n) to D by means of RSVP-TE signaling with the given ERO(n). Once the tunnel setup, it chooses a free label for the Stitching Label SL(n) and add a new entry in its MPLS LFIB with this SL(n) label. Then, it sends a PCRpt message to its PCE(n) with an RRO equal to {[LK(n), SL(n)], RRO(n)} and PLSP-ID(n). Once PCE(n) receives the PCRpt from the PCC BNe(n) with the RRO, PLSP-ID and LSP-TYPE = TBD2, it sends to the PCE(n-1) a PCRpt containing the RRO equal to {[LK(n), SL(n)]} and PLSP-ID(n). PCE(n) MAY adds {BN(n), D} in the RRO as loose path.

Steps i: Actions performed by all intermediate domains(i), for i = 2 to n-1

1. When the PCE(i) receives a PCRpt message from domain(i+1) with LSP-TYPE = TBD1, RRO = {[LK(i+1), SL(i+1)]} and PLSP-ID(i+1), it retrieves the ERO from the PKS(i) if necessary and sends to the PCC BNe(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, LSP-TYPE = TBD2 and End-Points Object = {BNe(i), BNo(i)} in order to inform the PCC BNe(i) that this local LSP tunnel(i) is part of an inter-domain path.
2. When the PCC BNe(i) received the PCInitiate message from its PCE(i), it setup the local LSP tunnels from BNe(i) to BNo(i) by means of RSVP-TE signaling with the given ERO(i).
3. When the BNo(i) receives an RSVP-TE Path message with an ERO = {x-1, [LK(i+1), SL(i+1)]} and End-Points Object = {BNe(i), BNo(i)}, it MUST install in its MPLS LFIB the SWAP instruction to label SL(i+1) with forward to LK(i+1) instead of the standard POP instruction.
4. Once the tunnel setup, it chooses a free label for the Stitching Label SL(i) and add a new entry in its MPLS LFIB with this SL(i) label. Then, it sends a PCRpt message to its PCE(i) with an RRO equal to {[LK(i), SL(i)]}, RRO(i)} and PLSP-ID(i).
5. Once PCE(i) receives the PCRpt from the PCC BNe(i) with the RRO and LSP-TYPE = TBD2, it sends to the PCE(i-1) a PCRpt message containing the RRO equal to {[LK(i), SL(i)]} and the PLSP-ID(i). PCE(i) MAY adds {BNe(i), BNo(i)} in the RRO as loose path.

Steps n: Actions performed at the source domain(1)

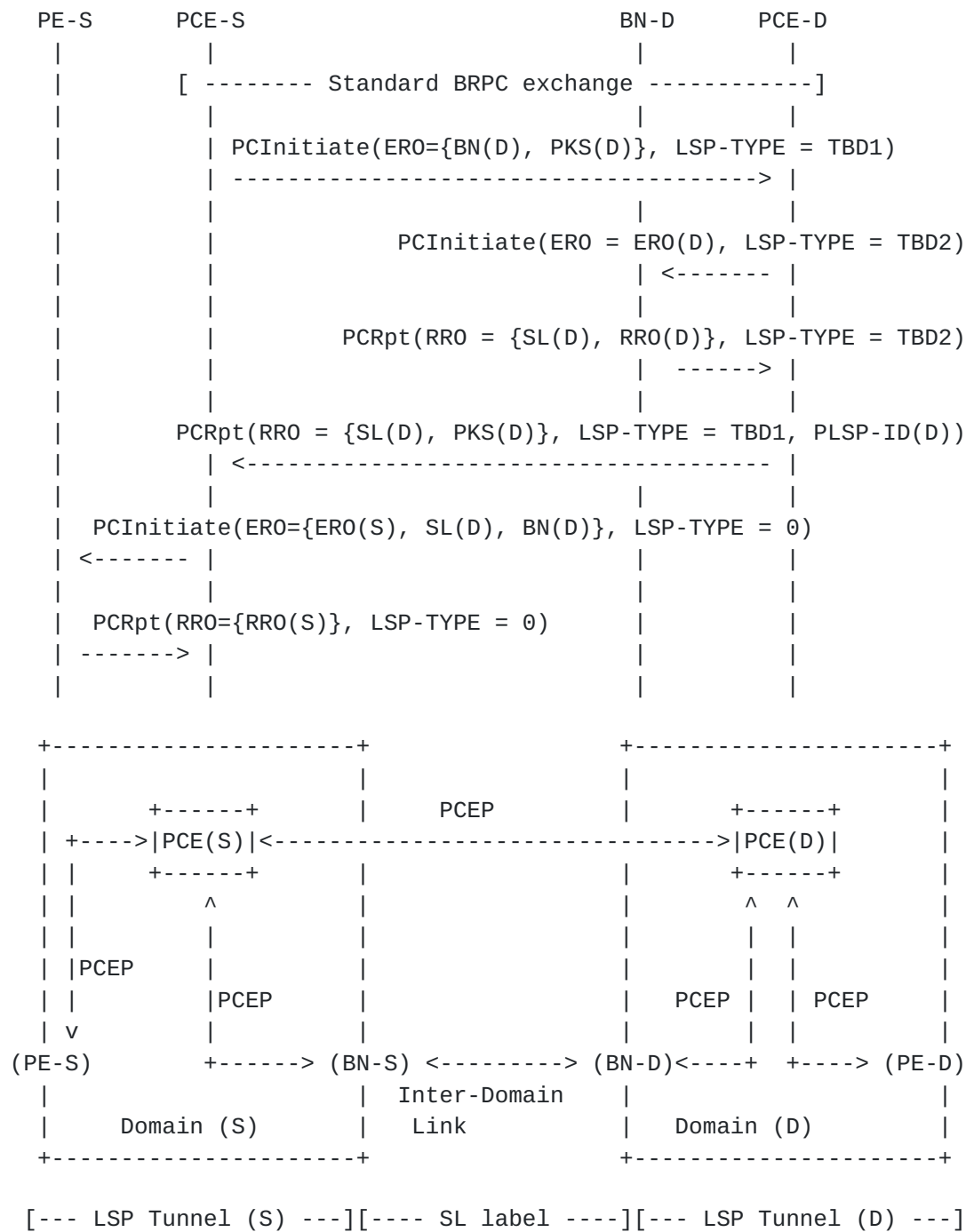
Once PCE(1) received the PCRpt message from PCE(2) with the RRO containing the label SL(2), it sends a PCInitiate message to PCC node S with ERO equal to {ERO(1), [LK(2), SL(2)]}, LSP_TYPE = 0 and End-Points Object = {S, BNo(1)}. This time, the LSP_TYPE is equal to 0 as the PCC S does not need to return a Stitching Label SL i.e. it is the head-end of the inter-domain path. Standard PCRpt message is sent back to PCE(1) by the PCC node S.

3.2. Example

In the figure below, two different domains S and D are interconnected through BN respectively BN-S and BN-D. PE-S and PE-D are edge routers. All routers in the figure are connected to their respective PCE through PCEP protocol. In this example, PCE(S) would setup an inter-domain path between PE-S and PE-D acting as source and destination of the tunnel. Intermediate routers between (PE-S, BN-S), (BN-D and PE-D) as well as RSVP-TE messages are not represented

to simplify the figure. But they are all presents. The following notation is used in the figure:

- o $PKS(D)$ = Path Key corresponding to the path from $BN(D)$ to $PE-D$
- o $ERO(D)$ = Explicit Route Object corresponding to the path from $BN(D)$ to $PE-D$ retrieves from $PKS(D)$
- o $RRO(D)$ = Record Route Object of local LSP tunnel(D) from $BN(D)$ to $PE-D$
- o $SL(D)$ = Stitching Label for local LSP tunnel from $BN(D)$ to $PE-D$
- o $ERO(S)$ = Explicit Route Object corresponding to the path from $PE-S$ to $BN(S)$
- o $RRO(S)$ = Record Route Object of local LSP tunnel(S) from $PE-S$ to $BN(S)$



Example of inter-domain path setup between two domains

3.3. Inter-domain LSP setup procedure completion failure

In case of error during LSP setup, PCRpt and or PCErrror messages MUST be used to signal the problem to the neighbor PCE domain backward. In particular, if new LSP-TYPE values defined in this memo are not

supported by the neighbor PCE or the PCC, the PCE, receptively the PCC, MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its neighbor PCE. If a PCE(i) receives a PCInitiate message from its peer PCE(i-1) without LSP-TYPE set to TBD1 or LSP-TYPE set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its peer PCE(i-1).

If a PCC or a PCE don't return an RRO or an RRO without the Stitching Label SL with the IP address of the associated link following a PCInitiate message with LSP-TYPE set to TBD1, the PCE MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = TBD4 (No Mandatory Stitching Label is present in the RRO).

In case of completion failure, the PCE(i) MUST propagate the PCErr message up to the PCE(1). In turn, PCE(1) MUST send a PCInitiate message (R flag set in the SRP Object as per draft pce initiated lsp [[I-D.ietf-pce-pce-initiated-lsp](#)]) to delete this inter-domain path to its neighbor PCEs. PCE(i) MUST propagate the PCInitiate message and remove their local LSP tunnel by means of PCInitiate message to their PCC BNe(i) and send back PCRpt message to PCE(i-1).

In case of error in domain(i+1), PCE(i) MAY add the AS number of domain(i+1) in the RRO to identify the faulty domain.

4. Hierarchical PCInitiate procedure

This section describes how to setup inter-domain paths than cross several different domains by using a Hierarchical method which is compatible to inter-domain path computation as describe in [RFC6805](#) [[RFC6805](#)].

4.1. Mode of operation

This section describes how PCInitiate and PCRpt messages are combined between PCE in order to setup inter-domain paths between a source domain(1) to a destination domain(n). S and D are respectively the source and destination of the inter-domain path. Domain(1) and domain(n) are different and connected through 0 or more intermediate domains denoted domain(i) with i = (2, n-1). Domains are directly connected when n = 2.

First, the Parent PCE contacts its Child PCE as per [RFC6805](#) [[RFC6805](#)] in order to compute the inter-domain path from S to D, where S and D are respectively a node in the domain(1) and domain(n). Path Key

confidentiality as per [RFC5520](#) [RFC5520] MAY be used to obfuscate the detailed ERO of the different domains(i). The resulting ERO is of the form (S, PKS(1), BNo(1), ..., BNe(i), PKS(i), BNo(i), ..., BNe(n), PKS(n), D).

The complete procedure follow the different steps described below:

Step 1: Initialization

Parent PCE sends a PCInitiate message to with an ERO = {PKS(n)} and End-Points = {BNe(n), D}. Then, PCE(n) retrieves the ERO from the PKS(n) if necessary and sends to BNe(n) a PCInitiate message with the ERO(n) = {BNe(n), ..., D}, LSP-TYPE= TBD2 and End-Points Object = {BN(n), D} in order to inform the PCC BNe(n) that this local LSP tunnel(n) is part of an inter-domain path. When the PCC BNe(n) received the PCInitiate message from its PCE(n), it setup the local LSP tunnel from entry BNe(n) to D by means of RSVP-TE signaling with the given ERO(n). Once the tunnel setup, it chooses a free label for the Stitching Label SL(n) and add a new entry in its MPLS LFIB with this SL(n) label. Then, it sends a PCRpt message to its PCE(n) with an RRO equal to {[LK(n), SL(n)], RRO(n)} and PLSP-ID(n). Once PCE(n) receives the PCRpt from the PCC BNe(n) with the RRO, PLSP-ID and LSP-TYPE = TBD2, it sends to its Parent PCE a PCRpt containing the RRO equal to {[LK(n), SL(n)]} and PLSP-ID(n). PCE(n) MAY adds {BN(n), D} in the RRO as loose path.

Steps i: Actions performed for all intermediate domains(i), for i = n-1 to 2

1. Parent PCE sends a PCInitiate message to Child PCE(i) with LSP-TYPE = TBD1, ERO = {PKS(i), [LK(i+1), SL(i+1)]} and End-Points = {BNe(i), BNo(i)}
2. Then, PCE(i) retrieves the ERO from the PKS(i) if necessary and sends to the PCC BNe(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, LSP-TYPE = TBD2 and End-Points Object = {BNe(i), BNo(i)} in order to inform the PCC BNe(i) that this local LSP tunnel(i) is part of an inter-domain path.
3. When the PCC BNe(i) received the PCInitiate message from its PCE(i), it setup the local LSP tunnel from BNe(i) to BNo(i) by means of RSVP-TE signaling with the given ERO(i).
4. When the BNo(i) receives an RSVP-TE Path message with an ERO = {x-1, [LK(i+1), SL(i+1)]} and End-Points Object = {BNe(i), BNo(i)}, it MUST install in its MPLS LFIB the SWAP instruction to label SL(i+1) with forward to LK(i+1) instead of the standard POP instruction.

5. Once the tunnel setup, it chooses a free label for the Stitching Label $SL(i)$ and add a new entry in its MPLS $L(F)IB$ with this $SL(i)$ label. Then, PCC $BNe(i)$ sends a PCRpt message to its PCE(i) with an RRO equal to $\{[LK(i), SL(i)], RRO(i)\}$ and PLSP-ID(i).
6. Once PCE(i) receives the PCRpt from the PCC $BNe(i)$ with the RRO and LSP-TYPE = TBD2, it sends to its Parent PCE a PCRpt message containing the RRO equal to $\{[LK(i), SL(i)]\}$ and the PLSP-ID(i). PCE(i) MAY adds $\{BNe(i), BNo(i)\}$ in the RRO as loose path.
7. Once Parent PCE receives the PCRpt from the Child PCE(i), it stores the corresponding PLSP-ID for this inter-domain tunnel part

Steps n: Actions performed to the source domain(1)

Finally, Parent PCE sends a last PCInitiate message to Child PCE(1) with LSP-TYPE = TBD1, ERO = $\{PKS(1), [LK(2), SL(2)]\}$ and End-Points = $\{S, BNo(1)\}$. In turn, Child PCE(1) sends a PCInitiate message to PCC node S with ERO equal to $\{ERO(1), [LK(2), SL(2)]\}$, LSP_TYPE = 0 and End-Points Object = $\{S, BNo(1)\}$. This time, the LSP_TYPE is equal to 0 as the PCC S does not need to return a Stitching Label SL, i.e. it is the head-end of the inter-domain path. Standard PCRpt message is sent back to PCE(1) by the PCC node S. In turn, Child PCE(1) send a final PCRpt message to the Parent PCE with the PSLP-ID(1). PCE(1) MAY adds $\{S, BNo(1)\}$ in the RRO as loose path.

4.2. Inter-domain LSP setup procedure completion failure

In case of error during LSP setup, PCRpt and or PCErr messages MUST be used to signal the problem to the Parent PCE. In particular, if new LSP-TYPE values defined in this memo are not supported by the Child PCE or the PCC, the Child PCE, receptively the PCC, MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE. If Child PCE(i) receives a PCInitiate message from its Parent PCE without LSP-TYPE set to TBD1 or LSP-TYPE set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE.

If a Child PCE or a PCC don't return an RRO or an RRO without the Stitching Label SL with the IP address of the associated link following a PCInitiate message with LSP-TYPE set to TBD1, the Parent PCE, respectively the Child PCE, MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = TBD4 (No Mandatory Stitching Label is present in the RRO).

In case of completion failure, the Parent PCE MUST send a PCInitiate message (R flag set in the SRP Object as per draft pce initiated lsp [[I-D.ietf-pce-pce-initiated-lsp](#)]) to delete this inter-domain path to the Child PCEs that already setup their respective part of the inter-domain tunnel. Child PCE(i) MUST remove their local LSP tunnel by means of PCInitiate message with R flag set to 1 to their PCC BNe(i) and send back PCRpt message to the Parent PCE.

5. Inter-domain LSP Management

This section describe how inter-domain LSPs could be manage.

5.1. Identification of inter-domain tunnels

First, in order to manage inter-domain tunnels composed by the stitching or nesting of local tunnels, it is important to identify them. For this purpose, PLSP-ID managed by PCEs are combined to one provided by PCCs to form global identifier as follow:

- o PCE(i) in the Backward Recursive method or the Child PCE in Hierarchical method MUST create a new unique PLSP-ID for this inter-domain LSP part and MUST send it in the PCRpt message, to the PCE(i-1), respectively the Parent PCE. In addition this new PLSP-ID MUST be associated to the one received from the PCC that instantiate the local tunnel part for further reference.
- o In Hierarchical mode, Parent PCE MUST store and associate the different PLSP-ID(i)s received from the different Child PCE(i)s in order to identify the different part of the inter-domain paths.
- o In Backward Recursive method, PCE(i) MUST store and associate its PLSP-ID(i) and the PLSP-ID(i+1) it received from the PCE(i+1). PCE(n) i.e. the last one in the chain, don't need to perform such association.

Further reference to the inter-domain tunnel will use this PLSP-ID(i). In Backward Recursive method, PCE(i) MUST replace the PLSP-ID(i) by PLSP-ID(i+1) in the PCUpd message before propagating it to PCE(i+1) and PCE(i) MUST replace the PLSP-ID(i+1) by PLSP-ID(i) in the PCRpt message before propagating it to the PCE(i-1). In Hierarchical method, Parent PCE MUST use the corresponding PLSP-ID(i) of the Child PCE(i).

5.2. Inter-domain LSP management

For the Backward Recursive method, each domain manages their respective local LSP tunnel part of an inter-domain path independently of each other. In particular, Stitching Label(i) is

managed by domain(i) and is of interest of domain(i-1) only. Thus, Stitching Label SL(i) is not supposed to be propagated to other domains. The same behavior apply to PLSP-ID(i). In Hierarchical method, the Parent PCE MUST ensure the correct distribution of Stitching Label SL(i) to Child PCE(i-1). The PLSP-ID(i) is kept for the usage of the Parent PCE and thus is not propagated.

If a PCE(i) needs to modify its local LSP tunnel(i) with PCUpd message to the PCC BNe(i), once PCRpt message received by the PCC BNe(i), it MUST sends a new PCRpt message to its neighbor PCE(i-1) in Backward Recursive method, respectively to Parent PCE in Hierarchical method, to advertise it of the modification. In particular. In this case PLSP-ID(i) is used to identify the inter-domain tunnel. PCE(i-1), respectively the Parent PCE, MUST propagate the PCRpt message if the modification imply the previous domain e.g. if the PCRpt indicates that the Stitching Label SL(i) has changed.

PCE(1), respectively Parent PCE, could modify the inter-domain path. For that purpose, it MUST sends a PCUpd message to its neighbor PCEs, respectively Child PCE, using the PLSP-ID it received. Each PCE(i) MUST process PCUpd message the same way they process PCInitiate message as define in [section 3.1](#) for Backward Recursive method and in [section 4.1](#) for Hierarchical method.

In case a failure appear in domain(i), e.g. tunnel becoming down, PCE(i) MUST sends a PCRpt message to its neighbor PCE(i-1), respectively its Parent PCE to advertise it of the problem in its local part of the inter-domain path. Once PCE(1), respectively Parent PCE, receives this PCRpt message indicating that the tunnel is down, it is up to the PCE(1), respectively Parent PCE to take appropriate correction e.g. start a new path computation to update the ERO.

5.3. Modification of inter-domain LSP

Modification of local LSP tunnel, BNe(i) and BNo(i) is left for further study.

5.4. Removal of inter-domain LSP

Deletion of inter-domain LSP is only possible by the inter-domain tunnel initiator. For Backward Recursive method, PCInitiate message with R flag set to 1 and PLSP-ID set accordingly to [section 5.1](#), is sent by PCE(1) to PCE(n) through PCE(i) and process the same way as describe in [section 3.1](#). For Hierarchical method, PCInitiate message with R flag set to 1 is sent by the Parent PCE to each Child PCE(i) with corresponding PLSP-ID(i) and process accordingly to [section 4.1](#).

6. Applicability

The newly introduced Stitching Label SL serves to stitch or nest part of local LSP tunnels to form an inter-domain path. Each domain is free to decide if the tunnel is stitched or nested. For example, a domain(i) may decide to nest the incoming local LSP tunnel into a higher hierarchy of tunnel for Traffic Engineering purpose. A PCE(i) may also decide to group local LSP tunnels part of inter-domain paths into a higher hierarchical tunnel to carry all these local LSP tunnels from one BNe(i) to one BNo(i).

To use Segment Routing instead of RSVP-TE to setup the local LSP tunnels as defined in draft pce segment routing [[I-D.ietf-pce-segment-routing](#)], PCE(i) MUST send PCInitiate message with LSP-TYPE = TBD3 instead of TBD2 to advertise their respective PCC that the local LSP tunnels is enforced by means of Segment Routing. Stitching Label SL(i) will be inserted in the label stack in order to become the top label in the stack when the packet reaches BNe(i+1): BNe(i) MUST install in its MPLS LFIB a SWAP instruction to replace the incoming Stitching Label SL(i) by the label stack given by the ERO(i) plus the Stitching Label SL(i+1) as the last label in the stack. The Stitching Label SL(i) serves as entry FEC for BNe(i) to identify the packets that follow the next Segment Path. When packet reaches BNo(i), the last label in the stack before the label SL(i+1) corresponds to the SID that designates BNe(i+1). This inter-domain SID could be obtained as per draft Egress Peer Engineering [[I-D.ietf-idr-bgpls-segment-routing-epe](#)].

The Stitching Label SL could serve to stitch Segment Path and RSVP-TE tunnel. Indeed, each domain is free to enforce its part of the inter-domain path with the underlying mechanism it chooses. Stitching Label SL will be part of the label stack in order to become the top label in the stack when reaching the BNe(i+1). This Stitching Label could be swapped as usual if the next domain uses RSVP-TE tunnel. When the previous domain uses a RSVP-TE tunnel, the Stitching Label will serve as key for the BNe(i+1) to determine which label stack it must push on top of the packet for a Segment Routing path.

During the instantiation procedure, if PCE(i) decides to reuse a local tunnel which is not yet part of an inter-domain tunnel, it SHOULD send a PCUpd message with LSP-TYPE = TBD2 to the PCC BNe(i) in order to request a Stitching Label SL(i) and new ERO(i) to include the Stitching Label SL(i+1) and the associated link to the previous ERO.

Inter-layer scenario is left for further study.

7. IANA Considerations

7.1. LSP-TYPE values

Draft pce lsp setup type [[I-D.ietf-pce-lsp-setup-type](#)] defines the PATH-SETUP-TYPE TLV and requests that IANA creates a registry to manage the value of the PATH_SETUP_TYPE TLV's PST field. IANA is requested to allocate a new code point in the PCEP PATH_SETUP_TYPE TLV PST field registry, as follows:

Value	Description	Reference
TBD1	Inter-Domain Traffic engineering end-to-end path is setup using Backward Recursive method	This Document
TBD2	Inter-Domain Traffic engineering local path is setup using RSVP-TE	This Document
TBD3	Inter-Domain Traffic engineering local path is setup using Segment Routing	This Document

7.2. PCEP-Error Object

IANA is requested to allocate code-points in the PCEP-ERROR Object Error Values registry for a new error-value or Error-Type 21 Invalid traffic engineering path setup:

Error-Value	Description
TBD4	Missing Mandatory Stitching Label in RRO

8. Security Considerations

No modification of PCE protocol (PCEP) has been requested by this draft which not introduce any issue regarding security. Concerning the PCEP session between PCEs, authors recommend to use the secure version of PCEP as defined in draft secure transport for PCEP [[RFC8253](#)] or use any other secure tunnel mechanism e.g. IPsec tunnel to transport PCEP session between PCE.

9. Acknowledgements

The authors want to thanks PCE's WG members.

10. Disclaimer

This work has been performed in the framework of the H2020-ICT-2014 project 5GEx (Grant Agreement no. 671636), which is partially funded by the European Commission. This information reflects the consortium's view, but neither the consortium nor the European Commission are liable for any use that may be done of the information contained therein.

11. References

11.1. Normative References

- [I-D.ietf-pce-lsp-setup-type]
Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying path setup type in PCEP messages", [draft-ietf-pce-lsp-setup-type-04](#) (work in progress), April 2017.
- [I-D.ietf-pce-pce-initiated-lsp]
Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", [draft-ietf-pce-pce-initiated-lsp-11](#) (work in progress), October 2017.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", [RFC 4655](#), DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", [RFC 5440](#), DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", [RFC 5441](#), DOI 10.17487/RFC5441, April 2009, <<https://www.rfc-editor.org/info/rfc5441>>.

- [RFC6805] King, D., Ed. and A. Farrel, Ed., "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", [RFC 6805](#), DOI 10.17487/RFC6805, November 2012, <<https://www.rfc-editor.org/info/rfc6805>>.
- [RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", [RFC 8231](#), DOI 10.17487/RFC8231, September 2017, <<https://www.rfc-editor.org/info/rfc8231>>.

11.2. Informative References

- [I-D.dong-pce-discovery-proto-bgp]
Dong, J., Chen, M., Dhody, D., Tantsura, J., Kumaki, K., and T. Murai, "BGP Extensions for Path Computation Element (PCE) Discovery", [draft-dong-pce-discovery-proto-bgp-07](#) (work in progress), July 2017.
- [I-D.ietf-idr-bgpls-segment-routing-epe]
Previdi, S., Filsfils, C., Patel, K., Ray, S., and J. Dong, "BGP-LS extensions for Segment Routing BGP Egress Peer Engineering", [draft-ietf-idr-bgpls-segment-routing-epe-13](#) (work in progress), June 2017.
- [I-D.ietf-pce-segment-routing]
Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "PCEP Extensions for Segment Routing", [draft-ietf-pce-segment-routing-10](#) (work in progress), October 2017.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", [RFC 3209](#), DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3473](#), DOI 10.17487/RFC3473, January 2003, <<https://www.rfc-editor.org/info/rfc3473>>.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", [RFC 4003](#), DOI 10.17487/RFC4003, February 2005, <<https://www.rfc-editor.org/info/rfc4003>>.

- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", [RFC 4206](#), DOI 10.17487/RFC4206, October 2005, <<https://www.rfc-editor.org/info/rfc4206>>.
- [RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", [RFC 5150](#), DOI 10.17487/RFC5150, February 2008, <<https://www.rfc-editor.org/info/rfc5150>>.
- [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](#), DOI 10.17487/RFC5520, April 2009, <<https://www.rfc-editor.org/info/rfc5520>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", [RFC 8253](#), DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.

Authors' Addresses

Olivier Dugeon
Orange
2, Avenue Pierre Marzin
Lannion 22307
France

Email: olivier.dugeon@orange.com

Julien Meuric
Orange
2, Avenue Pierre Marzin
Lannion 22307
France

Email: julien.meuric@orange.com

