Path Computation Element Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 23, 2021

O. Dugeon
J. Meuric
Orange Labs
Y. Lee
Huawei Technologies
D. Ceccarelli
Ericsson

January 19, 2021

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

Abstract

This document specifies how to combine a Backward Recursive or Hierarchical method with inter-domain paths in the context of stateful Path Computation Element (PCE). It relies on the PCInitiate message to set up independent paths per domain. Combining these different paths together enables to operate them as end-to-end inter-domain paths without the need for a signaling session between inter-domain border routers. A new Stitching Label is defined, new Path Setup Types, a new Association Type and a new PCEP communication Protocol (PCEP) Capability 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 July 23, 2021.

Table of Contents

1. Introduction

The PCE working group has produced a set of RFCs to standardize the behavior of the Path Computation Element as a tool to help MultiProtocol Label Switching - Traffic Engineering (MPLS-TE)/Generalized MPLS (GMPLS) Label Switched Paths (LSPs) and Segment Routing paths placement. This also includes the ability to compute inter-domain LSPs or Segment Routing paths following a distributed or hierarchical approach. To complement the original stateless mode, a stateful mode has been added and supports both passive and active control models. In particular, the new PCInitiate message allows a PCE to directly ask a PCC to set up an MPLS-TE/GMPLS LSP or a Segment Routing path. However, once computed, the inter-domain LSPs or Segment Routing paths are hard to set up in the underlying network. Especially, in operational networks, RSVP-TE signaling is usually not enabled between AS border routers. But, such RSVP-TE signaling is mandatory to set up contiguous LSP tunnels or to stitch or nest independent LSP tunnels to form the end-to-end inter-domain paths.

Looking at the different RFCs that describe the PCE architecture and in particular the PCE-based architecture [RFC4655], the PCE communication Protocol [RFC5440], BRPC [RFC5441] and H-PCE [RFC6805], the PCE is able to compute inter-domain paths, thus complementing the intra-domain computation. Such inter-domain paths could then serve as an Explicit Route Object (ERO) input for the RSVP-TE signaling to set up the tunnels within the underlying network. Three kinds of inter-domain paths could be established:

- Contiguous tunnel ([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.

- Stitching tunnel ([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 request the label that must be used by the previous domain to send the traffic in order that the MPLS packets follow the next LSP in the downstream domain. These labels are conveyed in the RSVP-TE Resv message.

- Nesting tunnel ([RFC4206]): This is similar to the stitching mode but, this time, with the possibility to set up tunnel hierarchy. For example, an LSP 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 cases, 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, typically only BGP sessions are set up 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 address some of the security concerns, 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.

For Segment Routing, issues are different as there is no signaling between routers. Here, the main problem comes from label stacking. The first issue concerns the size of the labels stack which is limited due to hardware constraint. The PCEP Extensions for Segment Routing [RFC8664] takes into account this limitation within the PCEP Capability when the PCEP session is established. Thus, taking into account Maximum Stack Depth (MSD), a PCE may be unable to find a solution when it computes an end-to-end inter-domain path. The second issue is related to the path confidentiality. With SR-TE, to express an explicit path, all Node-SID must be stacked by the head end router while some of the Node-SIDs are associated to routers of the next domains. It is clear that operators would not disclose details of their network, which includes Node-SIDs. Thus, it is not possible to stack remote labels for an end-to-end inter-domain path even if MSD constraint is respected.

The purpose of this memo is to take the benefit of active stateful PCE [RFC8231] and PCE-Initiated [RFC8281] modes to stitch or nest inter-domain paths directly using PCEP between domains' PCEs. This avoids using another signaling (e.g. RSVP-TE) at the inter-domain

border nodes, while keeping each operator free to independently set up their respective part of the inter-domain paths. The PCInitiate message is used in a Backward Recursive way like the PCReq message in BRPC [RFC5441], to recursively set up the end-to-end tunnel. PCRep message is used to automatically stitch or nest the different local LSPs. And, PCRep in conjunction with PCUpd messages are used to report, maintain, modify and remove inter-domain paths. This method is also applicable to Segment Routing to build inter-domain segment paths.

H-PCE [RFC6805] describes a Hierarchical PCE architecture which can be used for computing end-to-end paths for inter-domain MPLS-TE and GMPLS LSPs. Within this architecture, the Parent PCE (P-PCE) is used to compute a multi-domain path based on the domain connectivity information. A Child PCE (C-PCE) may be responsible for a single domain or multiple domains, it is used to compute the intra-domain path based on its domain topology information.

Stateful H-PCE [RFC8751] presents general considerations for stateful PCE(s) in the hierarchical PCE architecture. In particular, the behavior extends the existing stateful PCE mechanisms (including PCE-initiated LSP setup and active PCE usage) in the context of networks using the H-PCE architecture. Section 3.3.1 [RFC8751] describes the per-domain stitched LSP mode, where the individual per-domain LSPs are stitched together. PCInitiate message is also used to stitch the end-to-end tunnel. See section 4 for details.

1.1. General Assumptions

In the remainder of this document, the same references as per BRPC [RFC5441] are used and the following set of assumptions are made (see figure below):

- Domain refers to administrative partitions, i.e. an IGP area or an Autonomous System (AS).

- Inter-domain path is used to refer to a path that crosses two or more different domains as defined previously,

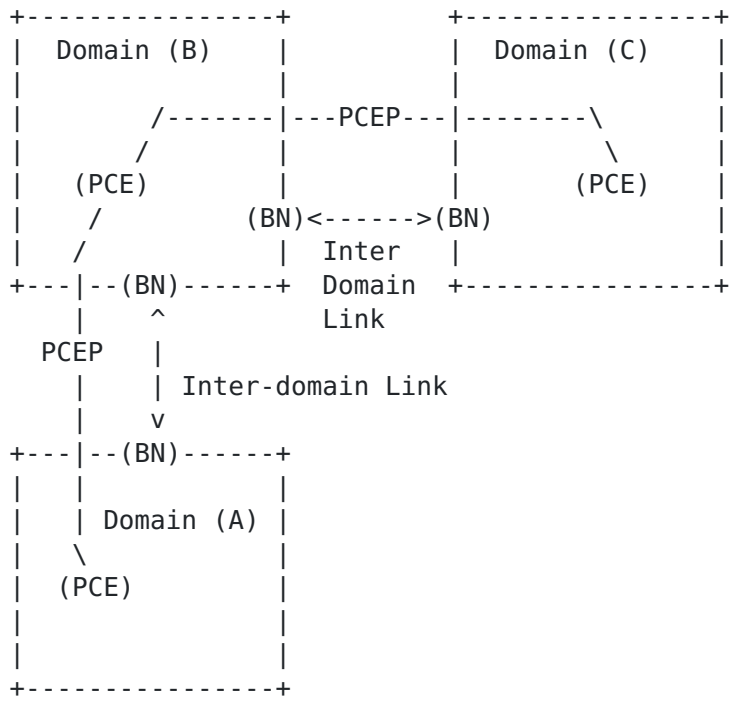
- At least one PCE is deployed in each domain. These PCEs are all active stateful-capable and can request to enforce LSPs in their respective domain by means of PCInitiate messages.

- LSRs, including border nodes, are PCC-enabled and support active stateful mode. PCEP sessions are established between these routers and their domains' PCE.

- Each PCE establishes a PCEP session with its respective neighbor domains' PCEs. The way a PCE discovers its neighboring PCEs is out of the scope of this document. This information could be administratively configured or automatically discovered through, for example,[I-D.dong-pce-discovery-proto-bgp].

- PCEs are able to compute an end-o-end path as per BRPC procedure [RFC5441] or as per H-PCE procedure (stateless [RFC6805] or stateful [RFC8751]).

- "Path" is a generic term to refer to both LSP 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).

AS: Autonomous System

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.

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

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

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

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

ERO(i): The Explicit Route Object scoped to domain(i)

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

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

LK(i): A Link that connect BN-ex(i-1) to BN-en(i). Note that BN-ex(i-1) could be connected to BN-en(i) by more than one link. LK(i) identifies which of the multiple links will be used for the inter-domain path setup. For inter-AS scenario, LK(i) represents the link between ASBR of domain i to the ASBR of domain i-1. For inter-area scenario, LK(i) is present only in IS-IS networks and represents the link between ABR of region L1, reciprocally L2, to the ABR of region L2, reciprocally L1.

Local path: A path that does not cross a domain border. It is set up either from entry BN-en, to output BN-ex or between both. This path could be enforce by means of RSVP-TE signaling or Segment Routing labels stack.

Local path(i): A Local path of domain(i)

PLSP-ID(i): A PLSP-ID that identifies, in the domain(i), the local part of an inter-domain path.

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 within the scope of domain(i).

PST: Path Setup Type

R(i,j): The router j of domain i

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

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

2. Stitching Label

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

2.1. Definition

The operation of stitch or nest a local path(i) to a local path(i+1) in order to form and inter-domain path mainly consists in defining the label that the output BN-ex(i) will use to send its traffic to the entry BN-en(i+1). Indeed, the entry BN-en(i+1) needs to identify the incoming traffic (e.g. IP packets), in order to know if this traffic must follow the local path(i+1) or not. Forwarding Equivalent Class (FEC) could be used for that purpose. But, when stitching or nesting tunnels, the FEC is reduced to the incoming label that the entry BN-en(i+1) has chosen for the local path(i+1).

In this memo, we introduce the term of "Stitching Label (SL)" to refer to this label. Such label is usually exchanged between output BN-ex(i) and entry BN-en(i+1) with the RSVP-TE signaling. But, as we want to avoid to use RSVP-TE signaling due to operational constraints, and allow compatibility support for Segment Routing, this Stitching Label is here conveyed by PCEP. In fact, the Explicit Route Object (ERO) and the Record Route Object (RRO) are already defined in order to transport (G)MPLS labels (for RSVP-TE or Segment Routing) in the PCEP signaling. Thus, the Stitching Label could be conveyed in the ERO and RRO without any modification of PCEP nor PCEP Objects.

As per RFC4003 [RFC4003], the Stitching Label will be conveyed as a companion of a link identifier (e.g. an IP address for numbered links). In our case, this is one of the endpoint IDs of the link LK(i) which connects BN-ex(i) to BN-en(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 endpoint IDs of the link LK(i) is used.

2.2. Inter-domain LSP-TYPE

Even if PCEP could convey the Stitching Label, a PCC is not aware that a PCE requests or provides such a label. For that purpose, this specification relies on the use of the PST as defined in [RFC8408] with new values (See IANA section of this memo) defined as follow:

- TBD1: Inter-Domain TE end-to-end path is set up using Backward Recursive or Hierarchical method. This new PST value MUST be set in a PCInitiate messages sends by a PCE(i-1) to its neighbor PCE(i) in the Backward Recursive method or by the Parent PCE to the Child PCE(i) to initiate a new inter-domain path. In its response, the neighbor PCE(1) or Child PCE(i) MUST return a Stitching Label SL with an identifier of the associated link in the RRO of the PCRpt message to PCE(i-1) or Parent PCE.

- TBD2: Inter-Domain TE local path is set up using RSVP-TE. This new PST 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 path(i) which is part of an inter-domain path. This PST value MUST be used by the PCE(i) only after receiving a PCInitiate message with an PST equal to TBD1 from a neighbor PCE(i-1) in the Backward Recursive method or Parent PCE in the Hierarchical method. In its response, the PCC of domain(i) MUST return a Stitching Label SL with the an identifier of associated link in the RRO of the PCRpt message.

- TBD3: Inter-Domain TE local path is set up using Segment Routing. This new PST 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 PST value MUST be used by the PCE(i) only after receiving a PCInitiate message with an PST equal to TBD1 from a neighbor PCE(i-1). In its response, the PCC MUST return a Stitching Label SL with an identifier of the associated link in the RRO of the PCRpt message.

3. Backward Recursive PCInitiate Procedure

This section describes how to set up inter-domain paths that cross different domains by using a Backward Recursive method. It is compatible with the 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 set up 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 (i.e. direct connection when n = 2) or more intermediate domains denoted domain(i) with i = [2, n-1].

First, the PCE(1) runs 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] SHOULD be used to obfuscate the detailed ERO(i) of the different domains(i). The resulting ERO is in the form {S, PKS(1), BN-ex(1), ..., BN-en(i), PKS(i), BN-ex(i), ..., BN-en(n), PKS(n), D} when Path Key is used and of the form {S, R(1,1), ..., R(1,k), BN-ex(1), ..., BN-en(i), R(i,1), ..., R(i,l), BN-ex(i), ..., BN-en(n), R(n,1), ..., R(n,m), D} otherwise . As subsequent domains are not aware about the computed end-to-end ERO in case of Virtual Source Path trees (VSPTs), the final ERO selected by the PCE(1) MUST be sent in the PCInitiate message to indicate to the subsequent PCEs which path has been finally chosen. PCE(1) MUST ensure that this ERO is self comprehensive by subsequent PCEs. Indeed, when a PCE(i) receives the ERO, it MUST be able to verify that this ERO matches its own scope and to determine the PCE(i+1). When Path Key is used, PCEs MUST encode the Path Key with a reachable IP address so that previous PCEs in the AS chain are able to join them. When Path Key is not used, the PCEs MUST be able to retrieve an IP address of the next PCE corresponding to the ERO (e.g., relying on a per prefix table).

The complete procedure with Path Key follows the different steps described below:

Steps 1: Initialization

Once ERO(S, D) is computed, PCE(1) sends a PCInitiate message to PCE(2) containing an ERO equal to {S, PKS(2), ..., PKS(i), ..., PKS(n), D}, PST = TBD1 and End-Points Object = (S, D). The ERO corresponds to the one PCE(1) has received from PCE(2) during the BRPC process in which only Path Key are kept. In case of multiple EROs, i.e. VSPT, 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 PST = TBD1 and ERO = {PKS(i), PKS(i+1), ..., PKS(n), D)}, it sends a

PCInitiate message to PCE(i+1) with a popped ERO and records its received PKS(i) part. All PCE(i)s generate the appropriate PCInitiate message to PCE(i+1) up to PCE(n), i.e. to the destination domain(n).

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

1. When a PCInitiate message reaches the destination domain(n), PCE(n) retrieves the ERO from the PKS(n) if necessary and sends to BN-en(n) a PCInitiate message with the ERO(n) = {BN-en(n), ..., D}, PST = TBD2 and End-Points Object = {BN(n), D} in order to inform the PCC BN-en(n) that this local path(n) is part of an inter-domain path.

2. When the PCC BN-en(n) receives the PCInitiate message from its PCE(n), it sets up the local path from entry BN-en(n) to D by means of RSVP-TE signaling with the given ERO(n).

3. Once the tunnel is set up, BN-en(n) chooses a free label for the Stitching Label SL(n) and adds a new entry in its MPLS L(F)IB 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).

4. Once PCE(n) receives the PCRpt from the PCC BN-en(n) with the RRO, PLSP-ID and PST = 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 add {PKS(n), D} in the RRO.

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 PST = TBD1, RRO = {[LK(i+1), SL(i+1)]} and PLSP-ID(i+1), it retrieves the ERO(i) from the PKS(i), recorded in step 1, and sends to the PCC BN-en(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, PST = TBD2 and End-Points Object = {BN-en(i), BN-ex(i)} in order to inform the PCC BN-en(i) that this local path(i) is part of an inter-domain path.

2. When the PCC BN-en(i) receives the PCInitiate message from its PCE(i), it sets up the local path from BN-en(i) to BN-ex(i) by means of RSVP-TE signaling with the given ERO(i).

3. Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is used to instruct the egress node of domain(i), i.e. BN-ex(i), to forward packets belonging to this tunnel with the Stitching Label. Both the Stitching Label and the identifier of the

interface are carried in the ERO = {..., [LK(i+1), SL(i+1)]} as the last SubObject in conformance to [RFC4003]. As a result, BN-ex(i) installs in its MPLS L(F)IB the SWAP instruction to label SL(i+1) with forward to LK(i+1).

4. Once the tunnel is set up, PCC BN-en(i) chooses a free label for the Stitching Label SL(i) and adds a new entry in its MPLS L(F)IB 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 BN-en(i) with the RRO and PST = TBD2, it sends to PCE(i-1) a PCRpt message containing the RRO equal to {[LK(i), SL(i)]} and the PLSP-ID(i). PCE(i) MAY add {PKS(i), ..., PKS(n)} in the RRO.

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

Once PCE(1) receives 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)]}, PST = 0 and End-Points Object = {S, BN-ex(1)}. This time, the PST is equal to 0 as the PCC S does not need to return a Stitching Label SL, because it is the head-end of the inter-domain path. A usual PCRpt message is sent back to PCE(1) by the PCC node S.

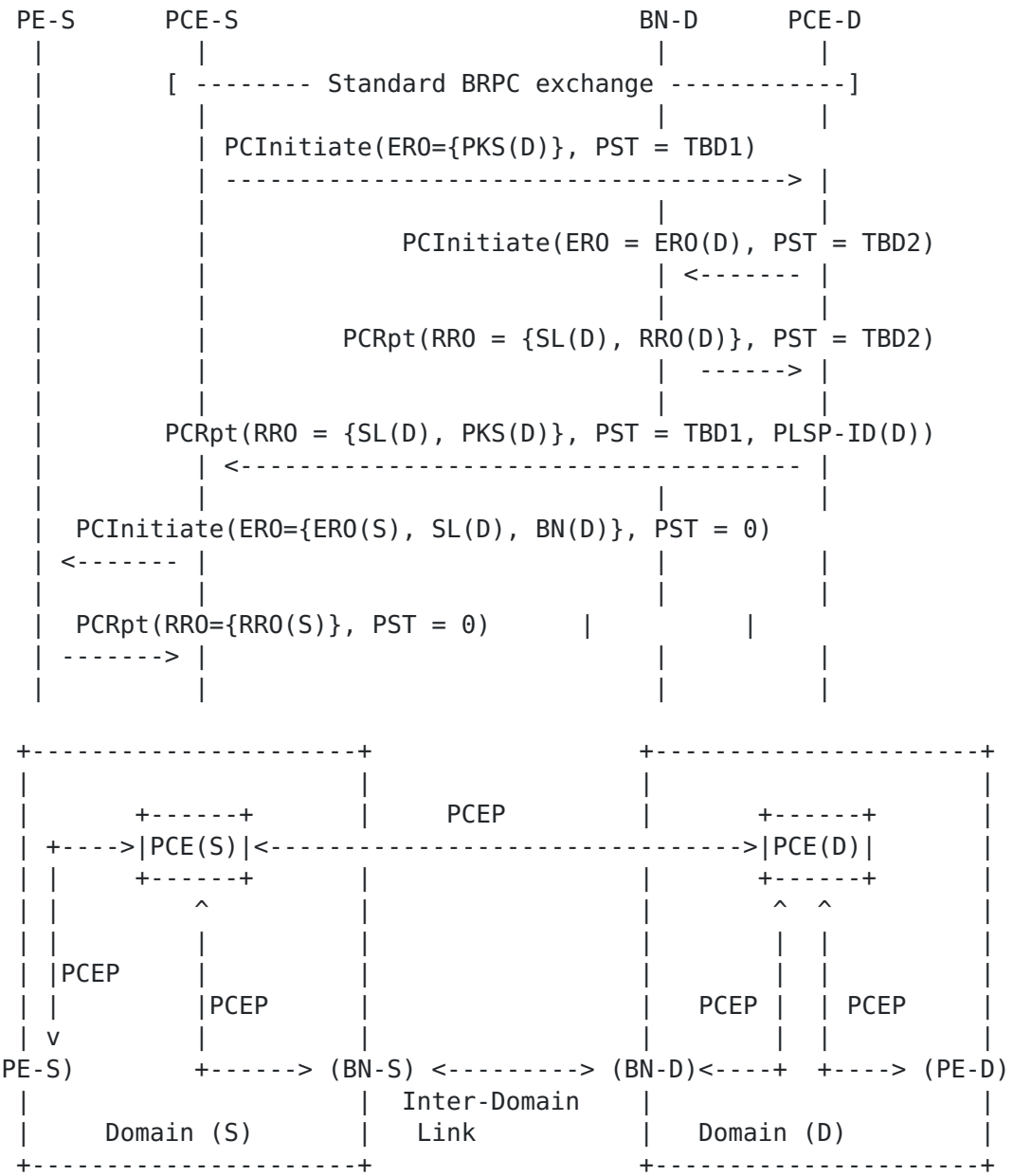
3.2. Example

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. In this example, we consider that PCE(S) needs to set up an inter-domain path between PE-S and PE-D acting as source and destination of the path. To simplify the figure, neither intermediate routers between (PE-S, BN-S), (BN-D and PE-D), nor RSVP-TE messages are represented, but they are all presents. The following notation is used (in this example, we use the PKS for the sake of simplicity):

- PKS(D) = Path Key corresponding to the path from BN(D) to PE-D
- ERO(D) = Explicit Route Object corresponding to the path from BN(D) to PE-D, retrieved from PKS(D)
- RRO(D) = Record Route Object of the local path(D) from BN(D) to PE-D
- SL(D) = Stitching Label for the local path from BN(D) to PE-D

- ERO(S) = Explicit Route Object corresponding to the path from PE-S to BN(S)

- RRO(S) = Record Route Object of local path(S) from PE-S to BN(S)



Example of inter-domain path setup between two domains

3.3. Completion Failure of Inter-domain Path Setup Procedure

In case of error during path setup, PCRpt and or PCErr messages MUST be used to signal the problem to the neighbor PCE domain backward. In particular, if the new PST values defined in this memo are not supported by the neighbor PCE or the PCC, the PCE, respectively the PCC, MUST return a PCErr message with Error-Type = 21 (TE 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 PST set to TBD1 or PST set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = 1 (Unsupported path setup type) to its peer PCE(i-1).

Following a PCInitiate message with PST set to TBD1, if a PCC or a PCE returns no RRO, or an RRO without the Stitching Label SL and an identifier of the associated link, the PCE MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = TBD5 (Mandatory Stitching Label missing 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 PCInitate message (R flag set in the SRP Object as per [RFC8281]) to tear down this inter-domain path from its neighbor PCEs. PCE(i) MUST propagate the PCInitiate message and remove its local path by means of PCInitiate message to its PCC BN-en(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 set up inter-domain paths that cross different domains by using a hierarchical method. It is compatible with inter-domain path computation as described in [RFC6805].

4.1. Mode of Operation

This section describes how PCInitiate and PCRpt messages are combined between PCEs in order to set up 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] 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] SHOULD be used to obfuscate the detailed ERO(i) of the different domains(i). The resulting ERO is of the form (S, PKS(1), BN-ex(1), ..., BN-en(i), PKS(i), BN-ex(i), ..., BN-en(n), PKS(n), D) when Path Key is used and of the form {S, R(1,1), ..., R(1,k), BN-ex(1), ..., BN-en(i), R(i,1), ..., R(i,l), BN-ex(i), ..., BN-en(n), R(n,1), ..., R(n,m), D} otherwise.

The complete procedure with Path Key follow the different steps described below:

Step 1: Initialization

1. The Parent PCE sends a PCInitiate message to Child PCE(n) with an ERO = {PKS(n)} and End-Points = {BN-en(n), D}. Then, PCE(n) retrieves the ERO from the PKS(n) (if necessary) and sends to BN-en(n) a PCInitiate message with the ERO(n) = {BN-en(n), ..., D}, PST = TBD2 and End-Points Object = {BN-en(n), D} in order to inform the PCC BN-en(n) that this local path(n) is part of an inter-domain path.

2. When the PCC BN-en(n) receives the PCInitiate message from its PCE(n), it sets up the local path from the entry BN-en(n) to D by means of RSVP-TE signaling with the given ERO(n).

3. Once the path is set up, it chooses a free label for the Stitching Label SL(n) and adds a new entry in its MPLS L(F)IB 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).

4. Once PCE(n) receives the PCRpt from the PCC BN-en(n) with the RRO, PLSP-ID and PST = 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 add PKS(n) in the RRO.

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

1. The Parent PCE sends a PCInitiate message to Child PCE(i) with PST = TBD1, ERO = {PKS(i), [LK(i+1), SL(i+1)]} and End-Points = {BN-en(i), BN-ex(i)}

2. Then, PCE(i) retrieves the ERO from the PKS(i) if necessary and sends to the PCC BN-en(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, PST = TBD2 and End-Points Object =

{BN-en(i), BN-ex(i)} in order to inform the PCC BN-en(i) that this local path(i) is part of an inter-domain path.

3. When the PCC BN-en(i) receives the PCInitiate message from its PCE(i), it sets up the local path from BN-en(i) to BN-ex(i) by means of RSVP-TE signaling with the given ERO(i).

4. Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is used to instruct the egress node of domain(i), i.e. BN-ex(i) to forward packets belonging to this tunnel with the Stitching Label. Both the Label Stitching and an identifier of the outgoing interface are carried in the ERO = {..., [LK(i+1), SL(i+1)]} as the last SubObject in conformance to [RFC4003]. So that, BN-ex(i) installs in its MPLS L(F)IB the SWAP instruction to label SL(i+1) with forward to LK(i+1) instead of the usual POP instruction.

5. Once the tunnel is set up, PCC BN-en(i) chooses a free label for the Stitching Label SL(i) and adds a new entry in its MPLS L(F)IB 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).

6. Once PCE(i) receives the PCRpt from the PCC BN-en(i) with the RRO and PST = 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 add PKS(i) in the RRO.

7. Once the Parent PCE receives the PCRpt from the Child PCE(i), it stores the corresponding PLSP-ID for this inter-domain path part.

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

Finally, the Parent PCE sends a last PCInitiate message to its Child PCE(1) with PST = TBD1, ERO = {PKS(1), [LK(2), SL(2)]} and End-Points = {S, BN-ex(1)}. In turn, Child PCE(1) sends a PCInitiate message to PCC node S with ERO equal to {ERO(1), [LK(2), SL(2)]}, PST = 0 and End-Points Object = {S, BN-ex(1)}. This time, the PST is equal to 0 as the PCC S does not need to return a Stitching Label SL, because it is the head-end of the inter-domain path. A usual PCRpt message is sent back to PCE(1) by the PCC node S. In turn, Child PCE(1) sends a final PCRpt message to the Parent PCE with the PSLP-ID(1). PCE(1) MAY add {S, BN-ex(1)} in the RRO as a loose path.

4.2. Completion Failure of Inter-domain Path Setup Procedure

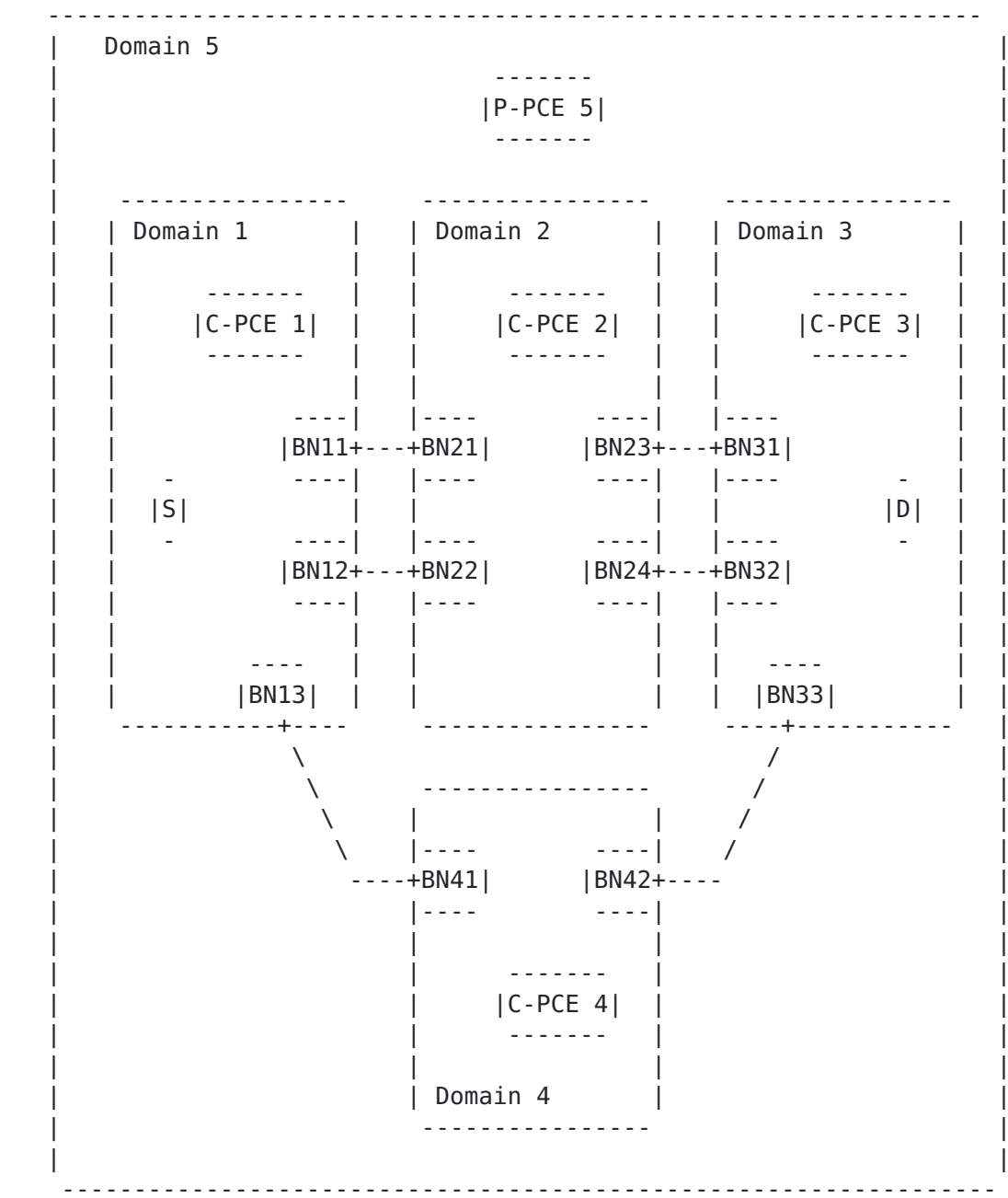
In case of error during path set up, PCRpt and or PCError messages MUST be used to signal the problem to the Parent PCE. In particular, if the new PST values defined in this memo are not supported by the Child PCE or the PCC, the Child PCE, respectively the PCC, MUST return a PCErr message with Error-Type = 21 (TE 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 PST set to TBD1 or PST set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE.

Following a PCInitiate message with PST set to TBD1, if a Child PCE or a PCC returns no RRO, or an RRO without the Stitching Label SL and an identifier of the associated link, the Parent PCE, respectively the Child PCE, MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = TBD5 (Mandatory Stitching Label missing in the RRO).

In case of completion failure, the Parent PCE MUST send a PCInitate message (R flag set in the SRP Object as per [RFC8281]) to tear down this inter-domain path from the Child PCEs that already set up their respective part of the inter-domain path. Child PCE(i) MUST remove its local path by means of PCInitiate message with R flag set to 1 to its PCC BN-en(i) and send back a PCRpt message to the Parent PCE.

4.3. Example for Stateful H-PCE Stiching Procedure

Taking the sample hierarchical domain topology example from [RFC6805] as the reference topology for the entirety of this section.



Hierarchical domain topology from RFC6805

Section 3.3.1 of [RFC8751] describes the per-domain stitched LSP mode and list all the steps needed. To support SL-based stitching, using the reference architecture described in the figure above, the steps are modified as follows (note that we do not use PKS in this example for simplicity):

Step 1: initialization

The P-PCE (PCE5) is requested to initiate a path. Steps 4 to 10 of section 4.6.2 of [RFC6805] are executed to determine the end-to-end path, which are split into per-domain paths, e.g. {S-BN41, BN41-BN33, BN33-D}.

Step 2: Path (BN33-D) at C-PCE3:

1. The P-PCE (P-PCE5) sends the initiate request to the C-PCE (C-PCE3) via PCInitiate message for path (BN33-D) with ERO={BN33..D} and PST = TBD1.

2. C-PCE3 further propagates the initiate message to BN33 with the ERO and PST = TBD2/TBD3 based on the setup type.

3. BN33 initiates the setup of the path and reports to the status ("GOING-UP") to C-PCE3.

4. C-PCE3 further reports the status of the path to the P-PCE (P-PCE5)

5. The node BN33 notifies the path state to C-PCE3 when the state is "UP"; it also sends the Stitching Label (SL33) in the RRO as {SL33,BN33..D}.

6. C-PCE3 further reports the status of the path to the P-PCE (P-PCE5) as well as sends the Stitching Label (SL33) in the RRO as {LK33,SL33,BN33..D}.

Step 3: Path (BN41-BN33) at C-PCE4

1. The P-PCE (P-PCE5) sends the initiate request to the C-PCE (C-PCE4) via PCInitiate message for path (BN41-BN33) with ERO={BN41..BN42,LK33,SL33,BN33} and PST = TBD1.

2. C-PCE4 further propagates the initiate message to BN41 with the ERO and PST = TBD2/TBD3 based on the setup type. In case of RSVP_TE, the node BN41 encode the Stitching Label SL33 as part of the ERO to make sure the node BN42 uses the label SL33 towards node BN33. In case of SR, the label SL33 is part of the label stack pushed at node BN41.

3. BN41 initiates the setup of the path and reports the path status ("GOING-UP") to C-PCE4.

4. C-PCE4 further reports the status of the path to the P-PCE (P-PCE5).

5. The node BN41 notifies the path state to C-PCE4 when the state is "UP"; it also sends the Stitching Label (SL41) in RRO as {LK41,SL41,BN41..BN33}.

6. C-PCE4 further reports the status of the to the P-PCE (P-PCE5) as well as sends the Stitching Label (SL41) in the RRO as {LK41,SL41,BN41..BN33}.

Step 3: Path (S-BN41) at C-PCE1

1. The P-PCE (P-PCE5) sends the initiate request to the C-PCE (C-PCE1) via PCInitiate message for path (S-BN41) with ERO={S..BN13,LK41,SL41,BN41}.

2. C-PCE1 further propagates the initiate message to node S with the ERO. In case of RSVP-TE, node S encodes the Stitching Label SL41 as part of the ERO to make sure the node BN13 uses the label SL41 towards node BN41. In case of SR, the label SL41 is part of the label stack pushed at node S.

3. S initiates the setup of the path and reports the path status ("GOING-UP") to C-PCE1.

4. C-PCE1 further reports the status of the path to the P-PCE (P-PCE5)

5. The node S notifies the path state to C-PCE1 when the state is "UP".

6. C-PCE1 further reports the status of the path to the P-PCE (P-PCE5).

In this way, per-domain paths are stitched together using the Stitching Label (SL). The per-domain paths MUST be set up from the destination domain towards the source domain one after the other.

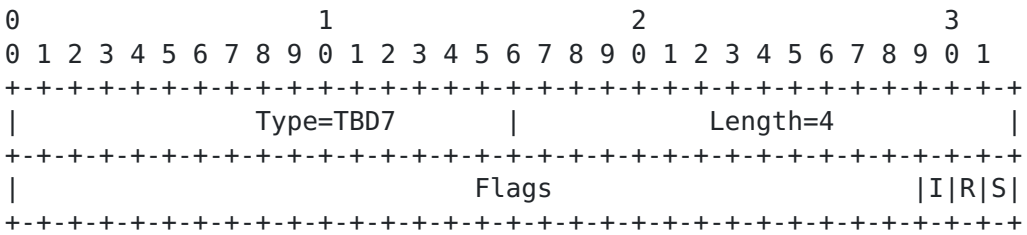
Once the per-domain path is set up, the entry BN chooses a free label for the Stitching Label SL and adds a new entry in its MPLS L(F)IB with this SL label. The SL from the destination domain is propagated to adjacent transit domain, towards the source domain at each step. This happens from the entry BN to C-PCE then to the P-PCE, and vice-versa. In case of RSVP-TE, the entry BN further propagates the SL label to the exit BN via RSVP-TE. In case of SR, the SL label is pushed as part of the SR label stack.

5. Inter-domain Path Management

This section describes how inter-domain paths could be managed.

5.1. Stitching Label PCE Capabilities

A PCE needs to know if its neighbor PCEs as well as PCCs are able to configure and provide a Stitching Label. The STITCHING-LABEL-PCE-CAPABILITY TLV is an optional TLV for use in the OPEN object for Stitching Label PCE capability advertisement. Its format is shown in the following figure:



STITCHING-LABEL-PCE-CAPABILITY TLV Format

The Type (16 bits) of the TLV is TBD7. The Length field is 16 bits long and has a fixed value of 4.

The value comprises a single 32 bits "Flags" field:

R (RSVP-TE-STITCHING-LABEL-CAPABILITY - 1 bit): if set to 1 by a PCC, the R flag indicates that the PCC is able to provide Stitching Labels, for RSVP-TE inter-domain paths, when requested by a PCE. If set to 1 by a PCE, the R flag indicates that the domain controlled by this PCE is able to set up inter-domain paths by means of RSVP-TE signaling.

S (SEGMENT-ROUTING-STITCHING-LABEL-CAPABILITY - 1 bit): if set to 1 by a PCC, the S flag indicates that the PCC is able to provide Stitching Labels, for Segment-Routing inter-domain paths, when requested by a PCE. If set to 1 by a PCE, the R flag indicates that the domain controlled by this PCE is able to set up inter-domain paths by means of Segment Routing.

I (INTER-DOMAIN-STITCHING-LABEL-CAPABILITY - 1 bit): if set to 1 by a PCE, the I flag indicates that the domain is supporting Stitching Label to set up inter-domain paths. This flag is reserved for PCEP session established between PCEs and MUST be kept unset by a PCC.

Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt.

PCCs MUST set the R and/or S flags and MUST NOT set the I flag when adding the Stitching Label Capability to the PCEP Open Message. The RSVP-TE-STITCHING-LABEL-CAPABILITY, respectively SEGMENT-ROUTING-STITCHING-LABEL-CAPABILITY, flag must be set by both the PCC and PCE in order to enable the configuration of Stitching Labels with RSVP-TE, respectively with Segment-Routing.

A PCE MUST set the I flag when establishing a PCEP session with a neighbor PCE when adding Stitching Label Capability to the PCEP Open Message. It MAY set R and/or S flags depending if the operator would like to keep confidential the technology used to set up inter-domain paths or not. The INTER-DOMAIN-STITCHING-LABEL-CAPABILITY flag must be set by both PCEs in order to enable inter-domain paths instantiation by means of Stitching Label.

5.2. Identification of Inter-domain Paths

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

- 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 path 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 instantiates the local path part for further reference.

- In the Hierarchical mode, the 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.

- In the 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, does not need to perform such association.

Further reference to the inter-domain path will use this PLSP-ID(i). In the Backward Recursive method, PCE(i) MUST replace the PLSP-ID(i) by PLSP-ID(i+1) in the PCUpd, PCRpt or PCinitiate 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 the Hierarchical method, the Parent PCE MUST use the corresponding PLSP-ID(i) of the Child PCE(i).

5.3. Inter-domain Association Group

In case of failure, a PCE(i) will received PCRpt messages from its PCCs and neighbors PCE(i+1) to synchronize the Inter-domain paths. In addition, it may received PCInitiate messages from its previous neighbors PCE(i-1) to re-initiate its inter-domain path part. As the PCE(i) may loose the PLSP-ID association, a new association group (within Association Object) is used to ease the association of the different parts of the inter-domain path: the local part and the PCE-to-PCE part. The use of the Association Object is MANDATORY in the Backward Recursive method and OPTIONAL in the Hierarchical method.

For that purpose, a new Inter-Domain Association Type with value TBD4 is defined. The first PCE in the Backward Recursive chain (the one which received the initial request) MUST send the PCInitiate message with an Association Object as follows:

- Association Type field MUST be set to new value TBD4
- Association ID MUST be set to a unique value. In case the Association ID field is too short or wraps, the first PCE MAY use the Extended Association ID to increase the number of association groups. The Association ID is managed locally by the PCE and does not need to be coordinated with neighbor or remote PCEs.
- IPV4 or IPv6 association source MUST be set to the IP address which identifies PCE(1) in domain(1).
- The Global Association Source TLV MUST be present and set with the ASN number of domain(1). It allows to create a globally unique association scope without putting constraint on operator's IP association source. Thus the IP Association Source is associated with the Global Association source to form a unique identifier.
- Extended Association ID MAY be present and MANDATORY if association ID is too short or wraps.

Subsequent PCE(i), for i = 2 to n, MUST send this Association Object as is to the local PCC and the neighbor PCE(i+1).

In case of error with the association group, a PCErr message MUST be raised with Error = 26 (Association Error) and Error value set accordingly. A new Error value TBD6 is defined to identify association of inter-domain paths.

In the Hierarchical method, the Parent PCE MAY act as the initiator of the Association and send to the Child PCEs an Association Object that follows the same rules as for the Backward Recursive method. In turn, Child PCEs MUST propagate the Association Object to the local PCCs as is.

5.4. Modification of Inter-domain Paths

For the Backward Recursive method, each domain manages their respective local path 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 the 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. Only the Association Object defined in section 5.2 is propagated if it is present.

If PCE(i) needs to modify its local path(i) with a PCUpd message to the PCC BN-en(i), once the PCRpt message received from the PCC BN-en(i), it MUST sends a new PCRpt message to advertise the modification. This message is targeted to its neighbor PCE(i-1) in the Backward Recursive method, respectively to the Parent PCE in the Hierarchical method. In this case PLSP-ID(i) is used to identify the inter-domain path. PCE(i-1), respectively the Parent PCE, MUST propagate the PCRpt message if the modification implies the upstream domain, e.g. if the PCRpt indicates that the Stitching Label SL(i) has changed.

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

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

5.5. Modification of Inter-domain Paths

Modification of local path, BN-en(i) and BN-ex(i) is left for further study.

5.6. Tear-Down of Inter-domain Paths

The tear-down of an inter-domain path is only possible by the inter-domain path initiator i.e. PCE(1). For the Backward Recursive method, a PCInitiate message with R flag set to 1, PLSP-ID set accordingly to section 5.1 and the Association Object with R flag set to 1, is sent by PCE(1) to PCE(n) through PCE(i), and processed the same way as described in section 3.1. For the Hierarchical method, a 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 processed according to section 4.1. Each domain PCE(i) is responsible to tear down its part of the path and the PCC MUST release both the Stitching label SL in its L(F)IB and the path when it receives the PCInitiate message with the R flag set to 1 and the corresponding PLSP-ID. The Association Group MUST also be removed by the PCC and PCE(i).

6. Applicability

The newly introduce Stitching Label SL serves to stitch or nest part of local paths to form an inter-domain path. Each domain is free to decide if the incoming path is stitched or nested and how the path is enforced, e.g. through RSVP-TE or Segment Routing. At the peering point, the Border Node BN-ex(i) MUST encapsulate the packet with the Stitching Label, i.e. the MPLS label prior to send them to the next Border Node BN-en(i+1). Thus, only RSVP-TE and Segment Routing over MPLS technology are detailed in the following sections.

6.1. RSVP-TE

In case of RSVP-TE, the Border Node BN-ex(i) needs to received the Stitching Label from BN-en(i) through the RSVP-TE message and install in its L(F)IB a SWAP instruction to the Stitching Label and forward it to the next Border Node BN-en(i+1). For that purpose, the Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is RECOMMENDED to instruct the Border Node BN-ex(i) of this action. Other mechanisms to program the L(F)IB could be used, e.g. NETCONF.

As the Stitching Label could serves to stitch or nest tunnels, a domain(i) may decide to nest the incoming LSPs into a higher hierarchy of LSPs for a Traffic Engineering purpose. A PCE(i) may also decide to group local LSPs part of inter-domain paths into a higher hierarchical LSP to carry all these local paths from a BN-en(i) to a BN-ex(i).

6.2. Segment Routing

To use Segment Routing instead of RSVP-TE to set up the local LSP tunnels as defined in [RFC8664], PCE(i) MUST send a PCInitiate message with PST = TBD3 instead of TBD2 to advertise its respective PCC that the local path is enforce by means of Segment Routing.

The Stitching Label SL(i+1) will be inserted into the label stack in order to become the top label in the stack when the packet reaches BN-en(i+1). Thus, the Stitching Label SL(i+1) serves as a FEC entry for BN-en(i+1) to identify the packets that follow the next Segment Path. For that purpose, BN-en(i+1) MUST install in its MPLS L(F)IB an instruction to replace the incoming Stitching Label SL(i+1) by the label stack given by the ERO(i+1) plus the Stitching Label SL(i+2), if any.

When a packet reaches BN-ex(i), the last label in the stack before the label SL(i+1) corresponds to a SID that allows to reach BN-en(i+1). When there are multiple interfaces between Border Nodes, BN-ex(i) needs to know how to send the packets to BN-en(i+1). Similarly to the Egress Control mechanism used with RSVP-TE, it is RECOMMENDED to use the inter-domain SID defined as per draft Egress Peer Engineering [I-D.ietf-idr-bgpls-segment-routing-epe] for that purpose. The inter-domain SID is announced by BN-ex(i) to PCE(i) through BGP-LS for each interface that connect BN-ex(i) to neighbors BN-en(i+1). Thus, the label stack will end with {BN-ex(i) SID, Inter-Domain SID, SL(i+1)} and should be processed as follows:

- The penultimate router of domain(i) pops its node SID, and sends the packet to the next node designated by the top label in the label stack, i.e. the node SID of BN-ex(i) or the adjacency SID of the link between the router and BN-ex(i).

- BN-ex(i) pops its node SID or its adjacency SID and looks up the next label in the stack, i.e. the inter-domain SID which corresponds to the interface to BN-en(i+1). BN-ex(i) pops this inter-domain SID as well and sends the packet to BN-ex(i) through the corresponding interface.

- BN-en(i+1) looks up the top label which is the Stitching Label SL(i+1), pops it and replaces it by the sub-sequent label stack.

Other mechanisms, e.g. NETCONF, could be used to configure the inter-domain SID on exit Border Nodes.

6.3. Mixing Technologies

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 PST = TBD2 to the PCC BN-en(i), in order to request a Stitching Label SL(i), and new ERO(i) to add the Stitching Label SL(i+1) and the associated link to the previous ERO.

[RFC8453] describes framework for Abstraction and Control of TE Networks (ACTN), where each Physical Network Controller (PNC) is equivalent to C-PCE and the Multi-Domain Service Coordinator (MDSC) to the P-PCE. The per-domain stitched LSP as per the Hierarchical PCE architecture described in Section 3.3.1 and Section 4.1 of [RFC8751] is well suited for ACTN. The Stitching Label mechanism as described in this document is well suited for ACTN when per-domain LSPs need to be stitched to form an E2E tunnel or a VN Member. It is to be noted that certain VNs require isolation from other clients. The SL mechanism described in this document can be applicable to the VN isolation use-case by uniquely identifying the concatenated stitching labels across multi-domain only to a certain VN member or an E2E tunnel.

As each operator is free to enforce the tunnel with its technology choice, it is a local policy decision for PCE(i) to instantiate the local part of the end-to-end tunnel by either RSVP-TE or Segment Routing. Thus, the PST value (i.e. TBD2 or TBD3) used in the PCinitiate message sent by the PCE(i) to the local PCC is determined by the local policy. How the local policy decision is set in the PCE is out of the scope of this memo. This flexibility is allowed because the SL principle allows to mix (data plane) technologies between domains. For example, a domain(i) could use RSVP-TE while domain(i+1) uses SR. The SL could serve to stitch indifferently Segment Paths and RSVP-TE tunnels. Indeed, the SL will be part of the label stack in order to become the top label in the stack when reaching the BN-en(i+1). This SL could be swapped as usual if the next domain uses RSVP-TE tunnels. When the upstream domain uses an RSVP-TE tunnel, the SL will serve as a key for the BN-en(i+1) to determine which label stack it must use on top of the packet for a Segment Routing path.

6.4. Inter-Area

If use cases for inter-AS are easily identifiable, this is less evident for inter-area. However, two scenarios have been identified:

- Paths between levels for IS-IS networks.
- Reduction of labels stack depth for Segment Routing.

Thus, the SL could be used to stitch or nest independent tunnels deployed through different IS-IS levels, even if there are controlled by the same PCE. IS-IS levels are considered as domains but under the control of the same PCE. In this scenario, there is no exchange between PCEs (it remains internal and implementation matter) and new TLVs are only applicable between the PCE and PCCs. The PCE requests to the different PCCs it identifies (i.e. BNs of the different IS-IS levels) to set up SLs and propagated them.

In large-scale networks, MSD could constraints the path computation in the possibility of path selection i.e. explicit expression of a path could exceeded the MSD. The SL could be used to split a too long explicit path regarding the MSD constraints. In this scenario, there is also no communications between PCEs and new TLVs are only used between PCE and PCCs.

7. IANA Considerations

7.1. Path Setup Type Values

[RFC8408] defines the PATH-SETUP-TYPE TLV. IANA is requested to allocate new code points in the PCEP PATH-SETUP-TYPE TLV PST field registry, as follows:

Value	Description	Reference
TBD1	Inter-domain TE end-to-end path is set up using the Backward Recursive method	This Document
TBD2	Inter-domain TE local path is set up using RSVP-TE signaling	This Document
TBD3	Inter-domain TE local path is set up using Segment Routing	This Document

7.2. Association Type Value

PCE Association Group [RFC8697] defines the ASSOCIATION Object and requests that IANA creates a registry to manage the value of the Association Type value. IANA is requested to allocate a new code point in the PCEP ASSOCIATION GROUP TLV Association Type field registry, as follows:

Association Type	Description
TBD4	Inter-domain Association Group

7.3. PCEP Error Values

IANA is requested to allocate code-points in the PCEP-ERROR Object Error Values registry for a new error-value of Error-Type 21 Invalid TE path setup and new error-value of Error-Type 26 Association Error:

Error-Type	Error-Value	Description
21	TBD5	Mandatory Stitching Label missing in the RRO
26	TBD6	Error in association of Inter-domain LSPs

7.4. PCEP TLV Type Indicators

IANA is requested to allocate a new TLV Type Indicator for the "Stitching Label PCE Capability" within the "PCEP TLV Type Indicators" subregistry of the "Path Computation Element Protocol (PCEP) Numbers" registry:

Value	Description	Reference
TBD7	STITCHING-LABEL-PCE-CAPABILITY	This Document

7.5. Stitching Label PCE Capability

IANA is requested to allocate a new subregistry, named "STITCHING-LABEL-PCE-CAPABILITY TLV Flag Field", within the "Path Computation Element Protocol (PCEP) Numbers" registry, to manage the Flag field in the STITCHING-LABEL-PCE-CAPABILITY TLV of the PCEP OPEN object (class = 1). New values are assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

- Bit number (counting from bit 0 as the most significant bit)
- Capability description
- Defining RFC

Value	Description	Reference
31	RSVP-TE-STITCHING-CAPABILITY	This Document
30	SEGMENT-ROUTING-STITCHING-CAPABILITY	This Document
29	INTER-DOMAIN-STITCHING-CAPABILITY	This Document

8. Security Considerations

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

9. Acknowledgements

The authors want to thanks PCE's WG members, and in particular Dhruv Dhody who greatly contributed to the Hierarchical section of this document and Quan Xiong for his advice.

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

[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>.

[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>.

[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>.

[RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <https://www.rfc-editor.org/info/rfc8281>.

[RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <https://www.rfc-editor.org/info/rfc8408>.

[RFC8697] Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "Path Computation Element Communication Protocol (PCEP) Extensions for Establishing Relationships between Sets of Label Switched Paths (LSPs)", RFC 8697, DOI 10.17487/RFC8697, January 2020, <https://www.rfc-editor.org/info/rfc8697>.

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., Talaulikar, K., 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-19 (work in progress), May 2019.

[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>.

[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>.

[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>.

[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>.

[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>.

[RFC8453] Ceccarelli, D., Ed. and Y. Lee, Ed., "Framework for Abstraction and Control of TE Networks (ACTN)", RFC 8453, DOI 10.17487/RFC8453, August 2018, <https://www.rfc-editor.org/info/rfc8453>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[RFC8751] Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., and D. King, "Hierarchical Stateful Path Computation Element (PCE)", RFC 8751, DOI 10.17487/RFC8751, March 2020, <https://www.rfc-editor.org/info/rfc8751>.

Authors' Addresses

Olivier Dugeon
Orange Labs
2, Avenue Pierre Marzin
Lannion 22307
France

Email: olivier.dugeon@orange.com

Julien Meuric
Orange Labs
2, Avenue Pierre Marzin
Lannion 22307
France

Email: julien.meuric@orange.com

Young Lee
Huawei Technologies
5340 Legacy Drive, Building 3
Plano TX 75023
USA

Email: leeyoung@huwaei.com

Daniele Ceccarelli
Ericsson
Torshamnsgatan, 48
Stockholm
Sweden

Email: daniele.ceccarelli@ericsson.com