PCE Working Group D. Dhody
Internet-Draft Z. Li
Intended status: Standards Track Huawei Technologies
Expires: April 25, 2019 October 22, 2018

# PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for P2MP LSPs
draft-dhody-pce-pcep-extension-pce-controller-p2mp-00

## Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

The PCE has been identified as an appropriate technology for the determination of the paths of point- to-multipoint (P2MP) TE Label Switched Paths (LSPs).

PCE was developed to derive paths for MPLS P2MP LSPs, which are supplied to the head end (root) of the LSP using PCEP. PCEP has been proposed as a control protocol to allow the PCE to be fully enabled as a central controller.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the P2MP LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the P2MP path while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP protocol extensions for using the PCE as the central controller for P2MP TE LSP.

## Status of This Memo

## Copyright Notice



## Table of Contents

## 1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the path while leveraging the existing PCE technologies as much as possible.

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static P2P LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label-forwarding instructions to program and what resources to reserve. The PCE-based controller keeps a view of the network and determines the paths of the end-to-end LSPs, and the controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

[RFC4857] describes how to set up point-to-multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) for use in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks. The PCE has been identified as a suitable application for the computation of paths for P2MP TE LSPs ([RFC5671]). The extensions of PCEP to request path computation for P2MP TE LSPs are described in [RFC8306]. Further [I-D.ietf-pce-stateful-pce-p2mp] specify the extensions that are necessary in order for the deployment of stateful PCEs to support P2MP TE LSPs as well as the setup, maintenance and teardown of PCE-initiated P2MP LSPs under the stateful PCE model.

This document extends [I-D.zhao-pce-pcep-extension-for-pce-controller] to specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static P2MP LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path with an added functionality of a P2MP branch node. As per [RFC4875], a branch node is an LSR that replicates the incoming data on to one or more outgoing interfaces. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for P2MP in PCECC architecture.

### 1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 2. Terminology

Terminologies used in this document is same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

## 3. Basic PCECC Mode

As described in [I-D.zhao-pce-pcep-extension-for-pce-controller], in this mode LSPs are provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label forwarding instructions to program and what resources to reserve. The controller uses PCEP to communicate with each router along the path of the end-to-end LSP. Note that the PCE-based controller will take responsibility for managing some part of the MPLS label space for each of the routers that it controls, and may taker wider responsibility for partitioning the label space for each router and allocating different parts for different uses. This is also described in section 3.1.2. of [RFC8283]. For the purpose of this document, it is assumed that label range to be used by a PCE is known and set on both PCEP peers. A future extension could add this capability to advertise the range via possible PCEP extensions as well.

This document extends the functionality to include support for central control instruction for replication at the branch nodes.

The rest of processing is similar to the existing stateful PCE mechanism for P2MP.

## 4. Procedures for Using the PCE as the Central Controller (PCECC) for P2MP

### 4.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231] and extended for P2MP [I-D.ietf-pce-stateful-pce-p2mp]. PCE as a central controller (PCECC) reuses existing Active stateful PCE mechanism as much as possible to control the LSP.

[I-D.zhao-pce-pcep-extension-for-pce-controller] extends PCEP messages - PCRpt, PCInitiate, PCUpd message for the Central Controller's Instructions (CCI) (label forwarding instructions in the context of this document). This documents specify the procedure for aditional instruction for branch node needed for P2MP.

### 4.2. PCECC Capability Advertisement

As per [I-D.zhao-pce-pcep-extension-for-pce-controller], during PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this PST=PCECC included in the PST list.

[I-D.zhao-pce-pcep-extension-for-pce-controller] also defines the PCECC Capability sub-TLV. A new M-bit is added in PCECC-CAPABILITY TLV to indicate support for PCECC-P2MP. A PCC MUST set M-bit in PCECC-CAPABILITY TLV and include STATEFUL-PCE-CAPABILITY TLV with P2MP bits set ([I-D.ietf-pce-stateful-pce-p2mp]) in OPEN Object to support the PCECC P2MP extensions defined in this document. If M-bit is set in PCECC-CAPABILITY TLV and N-bit in STATEFUL-PCE-CAPABILITY TLV is not set in OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD (P2MP capability was not advertised) and terminate the session.

### 4.3. LSP Operations

The PCEP messages pertaining to PCECC MUST include PATH-SETUP-TYPE TLV [RFC8408] with PST=PCECC [I-D.zhao-pce-pcep-extension-for-pce-controller] in the SRP object to clearly identify the PCECC LSP is intended.

#### 4.3.1. Basic PCECC LSP Setup

In order to setup a P2MP LSP based on PCECC mechanism, a PCC MUST delegate the P2MP LSP by sending a PCRpt message with PST set for PCECC and D (Delegate) flag (see [I-D.ietf-pce-stateful-pce-p2mp]) set in the LSP object.

P2MP-LSP-IDENTIFIER TLV [I-D.ietf-pce-stateful-pce-p2mp] MUST be included for PCECC LSP, the tuple uniquely identifies the P2MP LSP in the network. As per [I-D.zhao-pce-pcep-extension-for-pce-controller], the LSP object is included in central controller's instructions (label download) to identify the PCECC LSP for this instruction.

When a PCE receives PCRpt message with D flags and PST Type set, it calculates the P2MP tree and assigns labels along the path; and set up the path by sending PCInitiate message to each node along the path of the LSP, similar to [I-D.zhao-pce-pcep-extension-for-pce-controller]. The new extention required is the instructions on the branch nodes for replications to more than one outgoing interfaces with respective labels. The rest of the operations remains same as [I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.ietf-pce-stateful-pce-p2mp].

#### 4.3.2. Central Control Instructions

The new central controller's instructions (CCI) for the label operations in PCEP is done via the PCInitiate message, by defining a new PCEP Objects for CCI operations. Local label range of each PCC is assumed to be known at both the PCC and the PCE.

##### 4.3.2.1. Label Download

In order to setup an LSP based on PCECC, the PCE sends a PCInitiate message to each node along the path to download the Label instruction as described in Section 4.3.1.

The CCI object MUST be included, along with the LSP object in the PCInitiate message. The LSP-IDENTIFIER TLV MUST be included in LSP object. The SPEAKER-ENTITY-ID TLV SHOULD be included in LSP object.

As described in [I-D.zhao-pce-pcep-extension-for-pce-controller], if a node (PCC) receives a PCInitiate message which includes a Label to download as part of CCI, that is out of the range set aside for the PCE, it send a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (Label out of range). If a PCC receives a PCInitiate message but failed to download the Label entry, it sends a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (instruction failed).

Consider the example in the [I-D.ietf-teas-pcecc-use-cases] -

```
                          +----------+
                          |    R1    | Root node of the multicast LSP
                          +----------+
                              |6000
                          +----------+
           Transit Node   |    R2    |
           branch         +----------+
                          *  |   *  *
                     9001*   |   *   *9002
                        *    |   *    *
           +-----------+     |   *     +-----------+
           |    R4     |     |   *     |    R5     | Transit Nodes
           +-----------+     |   *     +-----------+
                      *      |   *      *     +
                   9003*     |   *     *      +9004
                        *    |   *    *       +
                        +-----------+  +-----------+
                        |    R3     |  |    R6     | Leaf Node
                        +-----------+  +-----------+
                         9005|
                        +-----------+
                        |    R8     | Leaf Node
                        +-----------+
```

PCECC would provision each node along the path and assign incoming and outgoing labels from R1 to {R6, R8} with the path: {R1, 6000}, {6000, R2, {9001,9002}}, {9001, R4, 9003}, {9002, R5, 9004} {9003, R3, 9005}, {9004, R6}, {9005, R8}. The operations on all nodes except R2 are same as [I-D.zhao-pce-pcep-extension-for-pce-controller]. The branch node (R2) needs to be instructed to replicate two copies of the incoming packet, and sent towards R4 and R5 with 9001 and 9002 labels respectively). This done via including 3 instances of CCI objects in the PCEP messages, one for each label in the example, 6000 for incoming and 9001/9002 for outgoing (along with remote nexthop). The message and procedure remains exactly as [I-D.zhao-pce-pcep-extension-for-pce-controller] with only distinction that more than one outgoing CCI MAY be present for the P2MP LSP.

#### 4.3.2.2. Label Cleanup

In order to delete an P2MP LSP based on PCECC, the PCE sends a central controller instructions via a PCInitiate message to each node along the path of the LSP to cleanup the Label forwarding instruction as per [I-D.zhao-pce-pcep-extension-for-pce-controller]. In case of branch nodes all instances of CCIs needs to be present in the PCEP message.

#### 4.3.3. PCE Initiated PCECC LSP

The LSP Instantiation operation is same as defined in [RFC8281] and [I-D.ietf-pce-stateful-pce-p2mp].

In order to setup a P2MP PCE Initiated LSP based on the PCECC mechanism, a PCE sends PCInitiate message with Path Setup Type set for PCECC (see Section 5.2) to the Ingress PCC (root).

The Ingress PCC MUST also set D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in LSP object of PCRpt message. The PCC responds with first PCRpt message with the status as "GOING-UP" and assigned PLSP-ID.

As described in [I-D.zhao-pce-pcep-extension-for-pce-controller], the label forwarding instructions from PCECC are send after the initial PCInitiate and PCRpt exchange. This is done so that the PLSP-ID and other LSP identifiers can be obtained from the ingress and can be included in the label forwarding instruction in the next PCInitiate message. The rest of the PCECC LSP setup operations are same as those described in Section 4.3.1.

#### 4.3.4. PCECC LSP Update

In case of a modification of PCECC P2MP LSP with a new path, the procedure and instructions as described in [I-D.zhao-pce-pcep-extension-for-pce-controller] apply.

#### 4.3.5. Re Delegation and Cleanup

In case of a redelgation and cleanup of PCECC P2MP LSP, the procedure and instructions as described in [I-D.zhao-pce-pcep-extension-for-pce-controller] apply.

#### 4.3.6. Synchronization of Central Controllers Instructions

The procedure and instructions are as per [I-D.zhao-pce-pcep-extension-for-pce-controller].

#### 4.3.7. PCECC LSP State Report

An Ingress PCC MAY choose to apply any OAM mechanism to check the status of LSP in the Data plane and MAY further send its status in PCRpt message (as per [I-D.ietf-pce-stateful-pce-p2mp]) to the PCE.

# 5. PCEP Objects

## 5.1. OPEN Object

### 5.1.1. PCECC Capability sub-TLV

The PCECC-CAPABILITY sub-TLV is an optional TLV for use in the OPEN Object for PCECC capability advertisement in PATH-SETUP-TYPE-CAPABILITY TLV as specified in [I-D.zhao-pce-pcep-extension-for-pce-controller].

This document adds a new flag (M-bit) in PCECC-CAPABILITY sub-TLV to indicate the support for P2MP in PCECC. A PCC MUST set M-bit in PCECC-CAPABILITY sub-TLV and set the N (P2MP-CAPABILITY), M (P2MP-LSP-UPDATE-CAPABILITY), and P (P2MP-LSP-INSTANTIATION-CAPABILITY) (as per [I-D.ietf-pce-stateful-pce-p2mp]) in STATEFUL-PCE-CAPABILITY TLV [RFC8231] to support the PCECC P2MP extensions defined in this document. If M-bit is set in PCECC-CAPABILITY sub-TLV and the P2MP bits in STATEFUL-PCE-CAPABILITY TLV are not set in OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD(P2MP capability was not advertised) and terminate the session.

## 5.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; [I-D.zhao-pce-pcep-extension-for-pce-controller] defines a PST value for PCECC, which is also used for P2MP.

## 5.3. CCI Object

The Central Control Instructions (CCI) Object [I-D.zhao-pce-pcep-extension-for-pce-controller] is used by the PCE to specify the forwarding instructions (Label information in the context of this document) to the PCC, and MAY be carried within PCInitiate or PCRpt message for label download which defined Object Type 1 for MPLS Label, which is also used for P2MP. The address TLVs [I-D.zhao-pce-pcep-extension-for-pce-controller] associates the next-hop information in case of an outgoing label.

If a node (PCC) receives a PCInitiate/PCUpd message with more than one CCI with O-bit set for outgoiing label and the node does not support the P2MP branch/replication capability, it MUST respond with PCErr message with Error-Type=2(Capability not supported).

## 6. Security Considerations

The security considerations described in [RFC8231], [RFC8281], [I-D.ietf-pce-stateful-pce-p2mp], and [I-D.zhao-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

## 7. Manageability Considerations

### 7.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/ disable PCECC P2MP capability as a global configuration.

### 7.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC P2MP capability.

### 7.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

### 7.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

### 7.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

### 7.6. Impact On Network Operations

PCEP extensions defined in this document do not put new requirements on network operations.

# 8. IANA Considerations

## 8.1. PCECC-CAPABILITY TLV

[I-D.zhao-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY TLV Flag Field registry, as follows:

| Bit | Description | Reference |
| --- | --- | --- |
| TBD | M((PCECC-P2MP-CAPABILITY)) | This document |

## 8.2. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

| Error-Type | Meaning | |
| --- | --- | --- |
| 19 | Invalid operation. | |
| | Error-value = TBD : | P2MP capability was not advertised |

# 9. Acknowledgments

# 10. References

## 10.1. 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>.

[RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <https://www.rfc-editor.org/info/rfc7420>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[I-D.ietf-pce-stateful-pce-p2mp] Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Path Computation Element (PCE) Protocol Extensions for Stateful PCE usage for Point-to-Multipoint Traffic Engineering Label Switched Paths", draft-ietf-pce-stateful-pce-p2mp-07 (work in progress), April 2018.

[I-D.zhao-pce-pcep-extension-for-pce-controller] Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-zhao-pce-pcep-extension-for-pce-controller-08 (work in progress), June 2018.

## 10.2. Informative References

[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>.

[RFC4857] Fogelstroem, E., Jonsson, A., and C. Perkins, "Mobile IPv4 Regional Registration", RFC 4857, DOI 10.17487/RFC4857, June 2007, <https://www.rfc-editor.org/info/rfc4857>.

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <https://www.rfc-editor.org/info/rfc4875>.

[RFC5671] Yasukawa, S. and A. Farrel, Ed., "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", RFC 5671, DOI 10.17487/RFC5671, October 2009, <https://www.rfc-editor.org/info/rfc5671>.

[RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <https://www.rfc-editor.org/info/rfc7025>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <https://www.rfc-editor.org/info/rfc7399>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <https://www.rfc-editor.org/info/rfc7491>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 8306, DOI 10.17487/RFC8306, November 2017, <https://www.rfc-editor.org/info/rfc8306>.

[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>.

[I-D.ietf-teas-pcecc-use-cases]
Zhao, Q., Li, Z., Khasanov, B., Dhody, D., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-02 (work in progress), October 2018.

[I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-09 (work in progress), October 2018.

## Appendix A. Contributor Addresses

Udayasree Palle

EMail: udayasreereddy@gmail.com

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahendrasingh@huawei.com

## Authors' Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

PCE Working Group Z. Li
Internet-Draft S. Peng
Intended status: Standards Track X. Geng
Expires: May 5, 2021 Huawei Technologies
M. Negi
RtBrick Inc
November 1, 2020

# PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for P2MP LSPs
draft-dhody-pce-pcep-extension-pce-controller-p2mp-05

## Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

The PCE has been identified as an appropriate technology for the determination of the paths of point-to-multipoint (P2MP) TE Label Switched Paths (LSPs).

PCE was developed to derive paths for MPLS P2MP LSPs, which are supplied to the head end (root) of the LSP using PCEP. PCEP has been proposed as a control protocol to allow the PCE to be fully enabled as a central controller.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the P2MP LSP can be calculated/set up/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network device along the P2MP path, while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP extensions for using the PCE as the central controller for P2MP TE LSP.

## Status of This Memo

Table of Contents

## 1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload the path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the path while leveraging the existing PCE technologies as much as possible.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP extensions for using the PCE as the central controller for static P2P LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label-forwarding instructions to program and what resources to reserve. The PCE-based controller keeps a view of the network and determines the paths of the end-to-end LSPs, and the controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

[RFC4857] describes how to set up point-to-multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) for use in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks. The PCE has been identified as a suitable application for the computation of paths for P2MP TE LSPs ([RFC5671]). The extensions of PCEP to request path computation for P2MP TE LSPs are described in [RFC8306]. Further [RFC8623] specify the extensions that are necessary in order for the deployment of stateful PCEs to support P2MP TE LSPs as well as the setup, maintenance and teardown of PCE-initiated P2MP LSPs under the stateful PCE model.

This document extends [I-D.ietf-pce-pcep-extension-for-pce-controller] to specify the procedures and PCEP extensions for using the PCE as the central controller for static P2MP LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path with an added functionality of a P2MP branch node. As per [RFC4875], a branch node is an LSR that replicates the incoming data on to one or more outgoing interfaces. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for P2MP in PCECC architecture.

### 1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 2. Terminology

Terminologies used in this document is the same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

## 3. Basic PCECC Mode

As described in [I-D.ietf-pce-pcep-extension-for-pce-controller], in this mode LSPs are provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label forwarding instructions to program and what resources to reserve. The controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

Note that the PCE-based controller will take responsibility for managing some part of the MPLS label space for each of the routers that it controls, and may taker wider responsibility for partitioning the label space for each router and allocating different parts for different uses. This is also described in section 3.1.2. of [RFC8283]. For the purpose of this document, it is assumed that the label range to be used by a PCE is known and set on both PCEP peers. A future extension could add the capability to advertise the range via possible PCEP extensions as well (see [I-D.li-pce-controlled-id-space]). The rest of the processing is similar to the existing stateful PCE mechanism.

This document extends the functionality to include support for central control instruction for replication at the branch nodes.

The rest of the processing is similar to the existing stateful PCE mechanism for P2MP [RFC8623].

# 4. Procedures for Using the PCE as a Central Controller (PCECC) for P2MP

## 4.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231] and extended for P2MP [RFC8623]. PCE as a Central Controller (PCECC) reuses the existing active stateful PCE mechanism as much as possible to control the LSPs.

[I-D.ietf-pce-pcep-extension-for-pce-controller] extends PCEP messages - PCRpt, PCInitiate message for the Central Controller's Instructions (CCI) (label forwarding instructions in the context of this document). This document specify the procedure for additional instruction for branch node needed for P2MP.

## 4.2. PCECC Capability Advertisement

As per [I-D.ietf-pce-pcep-extension-for-pce-controller], during PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of the PCECC extensions by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this PST=TBD [I-D.ietf-pce-pcep-extension-for-pce-controller] included in the PST list.

[I-D.ietf-pce-pcep-extension-for-pce-controller] also defines the PCECC Capability sub-TLV. A new M-bit is added in the PCECC-CAPABILITY sub-TLV to indicate support for PCECC-P2MP. A PCC MUST set M-bit in the PCECC-CAPABILITY sub-TLV and include STATEFUL-PCE-CAPABILITY TLV with the P2MP bits set ([RFC8623]) in the OPEN Object to support the PCECC P2MP extensions defined in this document. If the M-bit is set in PCECC-CAPABILITY sub-TLV and N-bit in the STATEFUL-PCE-CAPABILITY TLV is not set in the OPEN Object, the PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD2 (P2MP capability was not advertised) and terminate the session.

The rest of the processing is as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

## 4.3. LSP Operations

The PCEP messages pertaining to a PCECC includes the PATH-SETUP-TYPE TLV [RFC8408] with PST=TBD [I-D.ietf-pce-pcep-extension-for-pce-controller] in the SRP object to identify the PCECC LSP is intended as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

### 4.3.1. PCE-Initiated PCECC LSP

The LSP Instantiation operation is the same as defined in [RFC8281] and [RFC8623].

In order to set up a PCE-Initiated P2MP LSP based on the PCECC mechanism, a PCE sends PCInitiate message with Path Setup Type set to TBD for PCECC ([I-D.ietf-pce-pcep-extension-for-pce-controller]) to the ingress PCC (root node).

P2MP-LSP-IDENTIFIER TLV [RFC8623] MUST be included for the PCECC P2MP LSPs, the tuple uniquely identifies the P2MP LSP in the network. As per [I-D.ietf-pce-pcep-extension-for-pce-controller], the LSP object is included in the central controller's instructions (label download) to identify the PCECC P2MP LSP for this instruction. The handling of PLSP-ID is as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

The ingress PCC MUST also set D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in the LSP object of the PCRpt message. The PCC responds with a PCRpt message with the status set to "GOING-UP" and carrying the assigned PLSP-ID. As per [I-D.ietf-pce-pcep-extension-for-pce-controller] when the PCE receives this PCRpt message with the PLSP-ID, it assigns labels along the path; and sets up the path by sending a PCInitiate message to each node along the path of the P2MP Tree as per the PCECC technique. The CC-ID uniquely identifies the central controller instruction within a PCEP session. Each PCC further responds with the PCRpt messages including the central controller instruction (CCI) and the LSP objects.

As described in [I-D.ietf-pce-pcep-extension-for-pce-controller], the label forwarding instructions from PCECC are sent after the initial PCInitiate and PCRpt exchange. This is done so that the PLSP-ID and other LSP identifiers can be obtained from the ingress and can be included in the label forwarding instruction in the next PCInitiate message.

### 4.3.2. PCC-Initiated PCECC LSP

In order to set up a P2MP LSP based on the PCECC mechanism where the LSP is configured at the PCC, a PCC MUST delegate the P2MP LSP by sending a PCRpt message with PST set for PCECC and D (Delegate) flag (see [RFC8623]) set in the LSP object.

When a PCE receives the initial PCRpt message with the D flags and PST Type set to TBD, it SHOULD calculate the P2MP tree and assigns labels along the P2MP tree; and set up the P2MP LSP by sending PCInitiate message to each node along the path of the P2MP LSP as per

[I-D.ietf-pce-pcep-extension-for-pce-controller]. The new extension required is the instructions on the branch nodes for replications to more than one outgoing interface with the respective label. The rest of the operations remains the same as [I-D.ietf-pce-pcep-extension-for-pce-controller] and [RFC8623].

#### 4.3.3. Central Control Instructions

The new central controller's instructions (CCI) for the label operations in PCEP is done via the PCInitiate message as described in [I-D.ietf-pce-pcep-extension-for-pce-controller], by defining a new PCEP Objects for CCI operations. The local label range of each PCC is assumed to be known by both the PCC and the PCE.

##### 4.3.3.1. Label Download CCI

In order to set up an LSP based on PCECC, the PCE sends a PCInitiate message to each node along the path to download the Label instruction as described in Section 4.3.1 and Section 4.3.2.

The CCI object MUST be included, along with the LSP object in the PCInitiate message. As per [I-D.ietf-pce-pcep-extension-for-pce-controller], there are at most 2 instances of CCI object in the PCInitiate message. For PCECC-P2MP operations, multiple instances of CCI object for out-labels is allowed. Similarly to acknowledge the central controller instructions, the PCRpt message allows multiple instances of CCI object for PCECC-P2MP operations.

The LSP-IDENTIFIER TLV MUST be included in the LSP object. The SPEAKER-ENTITY-ID TLV SHOULD be included in LSP object.

As described in [I-D.ietf-pce-pcep-extension-for-pce-controller], if a node (PCC) receives a PCInitiate message which includes a Label to download, as part of CCI, that is out of the range set aside for the PCE, it send a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (Label out of range) (defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]). If a PCC receives a PCInitiate message but fails to download the Label entry, it sends a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (Instruction failed) (defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]).

Consider the example in the [I-D.ietf-teas-pcecc-use-cases] -

```
                          +-----------+
                          |    R1     | Root node of the P2MP TE LSP
                          +-----------+
                              |6000
                          +-----------+
           Transit Node   |    R2     |
           branch         +-----------+
                          *  |   *  *
                     9001*   |   *   *9002
                        *    |   *    *
           +-----------+     |   *     +-----------+
           |    R4     |     |   *     |    R5     | Transit Nodes
           +-----------+     |   *     +-----------+
                      *      |   *      *     +
                   9003*     |   *     *      +9004
                        *    |   *    *       +
                        +-----------+  +-----------+
                        |    R3     |  |    R6     | Leaf Node
                        +-----------+  +-----------+
                         9005|
                        +-----------+
                        |    R8     | Leaf Node
                        +-----------+
```

PCECC would provision each node along the path and assign incoming and outgoing labels from R1 to {R6, R8} with the path: {R1, 6000}, {6000, R2, {9001,9002}}, {9001, R4, 9003}, {9002, R5, 9004} {9003, R3, 9005}, {9004, R6}, {9005, R8}. The operations on all nodes except R2 are same as [I-D.ietf-pce-pcep-extension-for-pce-controller]. The branch node (R2) needs to be instructed to replicate two copies of the incoming packet, and sent towards R4 and R5 with 9001 and 9002 labels respectively). This done via including 3 instances of CCI objects in the PCEP messages, one for each label in the example, 6000 for incoming and 9001/9002 for outgoing (along with remote nexthop). The message and procedure remains exactly as [I-D.ietf-pce-pcep-extension-for-pce-controller] with only distinction that more than one outgoing CCI MAY be present for the P2MP LSP.

4.3.3.2.  Label Clean up CCI

In order to delete a P2MP LSP based on PCECC, the PCE sends a central controller instructions via a PCInitiate message to each node along the path of the P2MP tree to clean up the Label forwarding instruction as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. In case of branch nodes, all instances of CCIs needs to be present in the PCEP message.

#### 4.3.4. PCECC LSP Update

In case of a modification of PCECC P2MP LSP with a new path, the procedure, and instructions as described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply.

#### 4.3.5. Re-Delegation and Clean up

In case of a re-delegation and clean up of PCECC P2MP LSP, the procedure, and instructions as described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply.

#### 4.3.6. Synchronization of Central Controllers Instructions

The procedure and instructions are as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

#### 4.3.7. PCECC LSP State Report

An ingress PCC MAY choose to apply any OAM mechanism to check the status of LSP in the Data plane and MAY further send its status in PCRpt message (as per [RFC8623]) to the PCE.

#### 4.3.8. PCC-Based Allocations

The PCE can request the PCC to allocate the label using the PCInitiate message. The procedure and instructions are as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

## 5. PCEP Messages

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the extension to PCInitiate and PCRpt message for PCECC. For P2P LSP, only two instances of CCI objects can be included. In the case of the P2MP LSP, multiple CCI objects are allowed. The message format and other procedures continue to apply.

## 6. PCEP Objects

### 6.1. OPEN Object

#### 6.1.1. PCECC Capability sub-TLV

The PCECC-CAPABILITY sub-TLV is an optional TLV for use in the OPEN Object for PCECC capability advertisement in PATH-SETUP-TYPE-CAPABILITY TLV as specified in [I-D.ietf-pce-pcep-extension-for-pce-controller].

This document adds a new flag (M-bit) in the PCECC-CAPABILITY sub-TLV to indicate the support for P2MP in PCECC.

M (PCECC-P2MP-CAPABILITY - 1 bit - TBD1): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable of PCECC-P2MP capability.

A PCC MUST set the M-bit in the PCECC-CAPABILITY sub-TLV and set the N (P2MP-CAPABILITY), the M (P2MP-LSP-UPDATE-CAPABILITY), and the P (P2MP-LSP-INSTANTIATION-CAPABILITY) bits (as per [RFC8623]) in the STATEFUL-PCE-CAPABILITY TLV [RFC8231] to support the PCECC-P2MP extensions defined in this document. If the M-bit is set in PCECC-CAPABILITY sub-TLV and the P2MP bits (in the STATEFUL-PCE-CAPABILITY TLV) are not set in the OPEN Object, a PCEP speaker SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD2 (P2MP capability was not advertised) and terminate the session.

## 6.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; [I-D.ietf-pce-pcep-extension-for-pce-controller] defines a PST value for PCECC, which is applicable for P2MP LSP as well.

## 6.3. CCI Object

The Central Control Instructions (CCI) Object [I-D.ietf-pce-pcep-extension-for-pce-controller] is used by the PCE to specify the forwarding instructions (Label information in the context of this document) to the PCC, and optionally carried within PCInitiate or PCRpt message for label download/report. The CCI Object Type 1 for MPLS Label is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller], which is used for the P2MP LSPs as well. The address TLVs are defined in [I-D.ietf-pce-pcep-extension-for-pce-controller], they associate the next-hop information in case of an outgoing label.

If a node (PCC) receives a PCInitiate message with more than one CCI with O-bit set for the outgoing label and the node does not support the P2MP branch/replication capability, it MUST respond with PCErr message with Error-Type=2 (Capability not supported) (defined in [RFC5440]).

The rest of the processing is same as [I-D.ietf-pce-pcep-extension-for-pce-controller].

## 7. Security Considerations

The security considerations described in [RFC8231], [RFC8281], [RFC8623], and [I-D.ietf-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

## 8. Manageability Considerations

### 8.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC-P2MP capability as a global configuration.

### 8.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC-P2MP capability.

### 8.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

### 8.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

### 8.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

## 8.6. Impact On Network Operations

PCEP extensions defined in this document do not put new requirements on network operations.

# 9. IANA Considerations

## 9.1. PCECC-CAPABILITY sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY sub-TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field registry, as follows:

| Bit | Description | Reference |
|---|---|---|
| TBD1 | P2MP | This document |

## 9.2. PCEP-Error Object

IANA is requested to allocate a new error value within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

```
Error-Type   Meaning
----------   -------
19           Invalid operation.

              Error-value = TBD2 :                P2MP capability was
                                                  not advertised
```

# 10. Acknowledgments

# 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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[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>.

[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>.

[RFC8623] Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Stateful Path Computation Element (PCE) Protocol Extensions for Usage with Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 8623, DOI 10.17487/RFC8623, June 2019, <https://www.rfc-editor.org/info/rfc8623>.

[I-D.ietf-pce-pcep-extension-for-pce-controller]
Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-ietf-pce-pcep-extension-for-pce-controller-07 (work in progress), September 2020.

## 11.2. Informative References

[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>.

[RFC4857] Fogelstroem, E., Jonsson, A., and C. Perkins, "Mobile IPv4 Regional Registration", RFC 4857, DOI 10.17487/RFC4857, June 2007, <https://www.rfc-editor.org/info/rfc4857>.

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <https://www.rfc-editor.org/info/rfc4875>.

[RFC5671] Yasukawa, S. and A. Farrel, Ed., "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", RFC 5671, DOI 10.17487/RFC5671, October 2009, <https://www.rfc-editor.org/info/rfc5671>.

[RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <https://www.rfc-editor.org/info/rfc7025>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <https://www.rfc-editor.org/info/rfc7399>.

[RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <https://www.rfc-editor.org/info/rfc7420>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <https://www.rfc-editor.org/info/rfc7491>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <https://www.rfc-editor.org/info/rfc7525>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 8306, DOI 10.17487/RFC8306, November 2017, <https://www.rfc-editor.org/info/rfc8306>.

[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>.

[I-D.ietf-teas-pcecc-use-cases]
Li, Z., Khasanov, B., Dhody, D., Zhao, Q., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-06 (work in progress), September 2020.

[I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-14 (work in progress), July 2020.

[I-D.li-pce-controlled-id-space]
Li, C., Chen, M., Wang, A., Cheng, W., and C. Zhou, "PCE Controlled ID Space", draft-li-pce-controlled-id-space-07 (work in progress), October 2020.

## Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

## Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: pengshuping@huawei.com

Xuesong Geng
Huawei Technologies
China

EMail: gengxuesong@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore, Karnataka  560102
India

EMail: mahend.ietf@gmail.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2019

D. Dhody
Z. Li
Huawei Technologies

October 22, 2018

# PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for SRv6

draft-dhody-pce-pcep-extension-pce-controller-srv6-00

## Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled (G)MPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. This document specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers for Segment Routing in IPv6 (SRv6), in addition to computing the SRv6 paths for packet flows and telling the edge routers what instructions to attach to packets as they enter the network.

## Status of This Memo

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 25, 2019.

Copyright Notice



Table of Contents

## 1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCECC architecture.

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment Routing (SR) technology leverages the source routing and tunneling paradigms. A source node can choose a path without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [RFC8402] provides an introduction to SR architecture. The corresponding IS-IS and OSPF extensions are specified in [I-D.ietf-isis-segment-routing-extensions] and [I-D.ietf-ospf-segment-routing-extensions] , respectively. It relies on a series of forwarding instructions being placed in the header of a packet. The list of segment forming the path is called the Segment List and is encoded in the packet header. Segment Routing can be applied to the IPv6 architecture with the Segment Routing Header (SRH) [I-D.ietf-6man-segment-routing-header]. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment. The segment routing architecture supports operations that can be used to steer packet flows in a network, thus providing a form of traffic engineering. [I-D.ietf-pce-segment-routing] and [I-D.negi-pce-segment-routing-ipv6] specify the SR specific PCEP extensions.

PCECC may further use PCEP protocol for SR SID (Segment Identifier) distribution on the SR nodes with some benefits. [I-D.zhao-pce-pcep-extension-pce-controller-sr] specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR SID distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network. This document extends this to include SRv6 SID distribution as well.

### 1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 2. Terminology

Terminologies used in this document is same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

## 3. PCECC SRv6

[I-D.ietf-pce-segment-routing] specifies extensions to PCEP that allow a stateful PCE to compute, update or initiate SR-TE paths for MPLS dataplane. An ingress node of an SR-TE path appends all outgoing packets with a list of MPLS labels (SIDs). This is encoded in SR-ERO subobject, capable of carrying a label (SID) as well as the identity of the node/adjacency label (SID). [I-D.negi-pce-segment-routing-ipv6] extends the procedure to include support for SRv6 paths.

As per [I-D.ietf-6man-segment-routing-header], an SRv6 Segment is a 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment". Further details are in an illustration provided in [I-D.filsfils-spring-srv6-network-programming]. The SR is applied to IPV6 forwarding plane using SRH. A SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool, or a PCE and provisioned on the ingress node. [I-D.negi-pce-segment-routing-ipv6] extended SR-ERO subobject capable of carrying an SRv6 SID as well as the identity of the node/adjacency represented by the SID.

As per [RFC8283], PCE as a central controller can allocate and provision the node/prefix/adjacency label (SID) via PCEP. As per [I-D.ietf-teas-pcecc-use-cases] this is also applicable to SRv6 SIDs.

Rest of the processing is similar to existing stateful PCE with SRv6 mechanism.

## 4. PCEP Requirements

Following key requirements for PCECC-SRv6 should be considered when‘ designing the PCECC based solution:

- o PCEP speaker supporting this draft MUST have the capability to advertise its PCECC-SRv6 capability to its peers.

- o PCEP speaker not supporting this draft MUST be able to reject PCECC-SRv6 related message with a reason code that indicates no support for it.

- o PCEP procedures MUST provide a means to update (or cleanup) the SRv6 SID to the PCC.

- o PCEP procedures SHOULD provide a means to synchronize the SRv6 SID allocations between PCE to PCC in the PCEP messages.

## 5. Procedures for Using the PCE as the Central Controller (PCECC) in SRv6

### 5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a central controller (PCECC) reuses existing Active stateful PCE mechanism as much as possible to control the LSP.

### 5.2. New Functions

This document uses the same PCEP messages and its extensions which are described in [I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr] for PCECC-SRv6 as well.

PCEP messages PCRpt, PCInitiate, PCUpd are also used to send LSP Reports, LSP setup and LSP update respectively. The extended PCInitiate message described in [I-D.zhao-pce-pcep-extension-for-pce-controller] is used to download or cleanup central controller’s instructions (CCIs) (SRv6 SID in scope of this document). The extended PCRpt message described in [I-D.zhao-pce-pcep-extension-for-pce-controller] is also used to report the CCIs (SRv6 SIDs) from PCC to PCE.

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify an object called CCI for the encoding of central controller’s instructions. [I-D.zhao-pce-pcep-extension-pce-controller-sr] extends the CCI by defining a object-type for segment routing. This document further extends the CCI by defining another object-type for SRv6.

### 5.3. PCECC Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions. A PCEP Speaker includes the "PCECC Capability" sub-TLV, described in [I-D.zhao-pce-pcep-extension-for-pce-controller].

A S-bit is added in PCECC-CAPABILITY sub-TLV to indicate support for PCECC-SR in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. This document adds another I-bit to indicate support for SR in IPv6. A PCC MUST set I-bit in PCECC-CAPABILITY sub-TLV and include SRv6-PCE-CAPABILITY sub-TLV ([I-D.negi-pce-segment-routing-ipv6]) in OPEN Object (inside the the PATH-SETUP-TYPE-CAPABILITY TLV) to support the PCECC SRv6 extensions defined in this document. If I-bit is set in PCECC-CAPABILITY sub-TLV and SRv6-PCE-CAPABILITY sub-TLV is not advertised in OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD(SRv6 capability was not advertised) and terminate the session.

### 5.4. PCEP session IP address and TEDB Router ID

As described in [I-D.zhao-pce-pcep-extension-pce-controller-sr], it is important to link the session IP address with the Router ID in TEDB for successful PCECC operations.

### 5.5. SRv6 Path Operations

The PCEP messages pertaining to PCECC-SRv6 MUST include PATH-SETUP-TYPE TLV [RFC8408] with PST=TBD in the SRP object to clearly identify the PCECC-SRv6 setup is intended.

#### 5.5.1. PCECC Segment Routing in IPv6 (SRv6)

Segment Routing (SR) as described in [RFC8402] depends on "segments" that are advertised by Interior Gateway Protocols (IGPs). The SR-node allocates and advertises the SID (node, adj etc) and flood via the IGP. This document proposes a new mechanism where PCE allocates the SRv6 SID centrally and uses PCEP to advertise the SRv6 SID. In some deployments PCE (and PCEP) are better suited than IGP because of centralized nature of PCE and direct TCP based PCEP session to the node.

##### 5.5.1.1. PCECC SRv6 Node/Prefix SID allocation

Each node (PCC) is allocated a node SRv6 SID by the PCECC. The PCECC sends PCInitiate message to update the SID table of each node. The TE router ID is determined from the TEDB or from "IPv4/IPv6 Router-ID" Sub-TLV [I-D.dhodylee-pce-pcep-ls], in the OPEN Object.

On receiving the SRv6 node SID allocation, each node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly. The PCInitiate message in this case MUST have FEC object.

On receiving the SRv6 node SID allocation:

For the local SID, node (PCC) needs to update SID with associated function (END function in this case) in "My Local SID Table" ([I-D.filsfils-spring-srv6-network-programming]).

For the non-local SID, node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly.

The PCInitiate message in this case MUST have FEC object.

The forwarding behavior and the end result is similar to IGP based "Node-SID" in SRv6. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node.

PCE relies on the Node/Prefix SRv6 SID cleanup using the same PCInitiate message.

#### 5.5.1.2. PCECC SRv6 Adjacency SID allocation

[I-D.ietf-pce-segment-routing] extends PCEP to allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

For PCECC SR, apart from node-SID, Adj-SID is used where each adjacency is allocated an Adj-SID by the PCECC. The PCECC sends PCInitiate message to update the label map of each Adj to the corresponding nodes in the domain. Each node (PCC) download the SRv6 SID instructions accordingly. Similar to SRv6 Node/Prefix Label allocation, the PCInitiate message in this case uses the FEC object.

The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SRv6.

The Path Setup Type for segment routing MUST be set for PCECC SRv6 = TBD (see Section 7.2). All PCEP procedures and mechanism are similar to [I-D.ietf-pce-segment-routing].

PCE relies on the Adj label cleanup using the same PCInitiate message.

#### 5.5.1.3. Redundant PCEs

[I-D.litkowski-pce-state-sync] describes synchronization mechanism between the stateful PCEs. The SRv6 SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC SRv6 state synchronization. Note that the SRv6 SIDs are independent to the PCECC-SRv6 paths, and remains intact till any topology change. The redundant PCEs MUST have a common view of all SRv6 SIDs allocated in the domain.

#### 5.5.1.4. Re Delegation and Cleanup

[I-D.zhao-pce-pcep-extension-for-pce-controller] describes the action needed for CCIs for the Basic PCECC LSP on this terminated session. Similarly actions should be applied for the SRv6 SID as well.

#### 5.5.1.5. Synchronization of SRv6 SID Allocations

[I-D.zhao-pce-pcep-extension-for-pce-controller] describes the synchronization of Central Controller's Instructions (CCI) via LSP state synchronization as described in [RFC8231] and [RFC8232]. Same procedures should be applied for SRv6 SIDs as well.

## 6. PCEP messages

The PCEP message is as per [I-D.zhao-pce-pcep-extension-pce-controller-sr].

## 7. PCEP Objects

### 7.1. OPEN Object

#### 7.1.1. PCECC Capability sub-TLV

[I-D.zhao-pce-pcep-extension-for-pce-controller] defined the PCECC-CAPABILITY TLV.

A new I-bit is defined in PCECC-CAPABILITY sub-TLV for PCECC-SRv6:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Type=TBD        |            Length=4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Flags                         |I|S|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

I (PCECC-SRv6-CAPABILITY - 1 bit): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable for PCECC-SRv6 capability and PCE would allocate node and Adj SRv6 SID on this session.

### 7.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]. PST = TBD is used when Path is setup via PCECC SRv6 mode.

On a PCRpt/PCUpd/PCInitiate message, the PST=TBD indicates that this path was setup via a PCECC-SRv6 based mechanism where either the SIDs were allocated/instructed by PCE via PCECC mechanism.

### 7.3. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the forwarding instructions is defined in [I-D.zhao-pce-pcep-extension-for-pce-controller]. This document defines another object-type for SRv6 purpose.

CCI Object-Type is TBD for SRv6 as below -

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            CC-ID                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      MT-ID    |   Algorithm   |          Flags      |N|E|V|L|O|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Reserved           |   SRv6 Endpoint Function      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                       SRv6 Identifier                         |
|                          (128-bit)                            |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                        Optional TLV                         //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The field CC-ID is as described in [I-D.zhao-pce-pcep-extension-for-pce-controller]. The field MT-ID, Algorithm, Flags are defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

Reserved: MUST be set to 0 while sending and ignored on receipt.

SRv6 Endpoint Function: 16 bit field representing supported functions associated with SRv6 SIDs.

SRv6 Identifier: 128 bit IPv6 addresses representing SRv6 segment.

[Editor's Note - It might be useful to seperate the LOC:FUNC part in the SRv6 SID]

### 7.4. FEC Object

The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object (and various Object-Types) are described in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. SRv6 Node SID MUST includes the FEC Object-Type 2 for IPv6 Node. SRv6 Adjacency SID MUST include the FEC Object-Type=4 for IPv6 adjacency. Further FEC object types would be added in future revisions.

## 8. Security Considerations

The security considerations described in [I-D.zhao-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

## 9. Manageability Considerations

### 9.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC SR capability as a global configuration.

### 9.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC SR capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC SR capability.

### 9.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

### 9.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

### 9.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

### 9.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCInitiate/PCUpd messages (as per [RFC8231]) sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

## 10. IANA Considerations

### 10.1. PCECC-CAPABILITY TLV

[I-D.zhao-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY TLV Flag Field registry, as follows:

| Bit | Description | Reference |
|---|---|---|
| TBD | I((PCECC-SRv6-CAPABILITY)) | This document |

### 10.2. New Path Setup Type Registry

IANA is requested to allocate new PST Field in PATH- SETUP-TYPE TLV. The allocation policy for this new registry should be by IETF Consensus. The new registry should contain the following value:

| Value | Description | Reference |
|---|---|---|
| TBD | Path is setup using PCECC-SRv6 mode | This document |

### 10.3. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

| Error-Type | Meaning | |
| --- | --- | --- |
| 19 | Invalid operation. | |
| | Error-value = TBD : | SRv6 capability was not advertised |

## 11. Acknowledgments

## 12. References

### 12.1. 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>.

[RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <https://www.rfc-editor.org/info/rfc7420>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[I-D.negi-pce-segment-routing-ipv6]
Negi, M., Dhody, D., Sivabalan, S., and P. Kaladharan, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-negi-pce-segment-routing-ipv6-03 (work in progress), October 2018.

[I-D.zhao-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-zhao-pce-pcep-extension-for-pce-controller-08 (work in progress), June 2018.

[I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-03 (work in progress), June 2018.

## 12.2. Informative References

[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>.

[RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <https://www.rfc-editor.org/info/rfc7025>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <https://www.rfc-editor.org/info/rfc7399>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <https://www.rfc-editor.org/info/rfc7491>.

[RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <https://www.rfc-editor.org/info/rfc8232>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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>.

[I-D.ietf-teas-pcecc-use-cases]
Zhao, Q., Li, Z., Khasanov, B., Dhody, D., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-02 (work in progress), October 2018.

[I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-09 (work in progress), October 2018.

[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-14 (work in progress), October 2018.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., Decraene, B., and J. Tantsura, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-19 (work in progress), July 2018.

[I-D.ietf-ospf-segment-routing-extensions]
Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", draft-ietf-ospf-segment-routing-extensions-25 (work in progress), April 2018.

[I-D.litkowski-pce-state-sync]
Litkowski, S., Sivabalan, S., and D. Dhody, "Inter Stateful Path Computation Element communication procedures", draft-litkowski-pce-state-sync-03 (work in progress), April 2018.

[I-D.dhodylee-pce-pcep-ls]
Dhody, D., Lee, Y., and D. Ceccarelli, "PCEP Extension for Distribution of Link-State and TE Information.", draft-dhodylee-pce-pcep-ls-11 (work in progress), June 2018.

[I-D.filsfils-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-filsfils-spring-srv6-network-programming-05 (work in progress), July 2018.

[I-D.ietf-6man-segment-routing-header]
Filsfils, C., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-14 (work in progress), June 2018.

## Appendix A. Contributor Addresses

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahendrasingh@huawei.com

## Authors' Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 5, 2021

Z. Li
S. Peng
X. Geng
Huawei Technologies
M. Negi
RtBrick Inc

November 1, 2020

# PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for SRv6

draft-dhody-pce-pcep-extension-pce-controller-srv6-05

## Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled (G)MPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. This document specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers for Segment Routing (SR) in IPv6 (SRv6), in addition to computing the SRv6 paths for packet flows and telling the edge routers what instructions to attach to packets as they enter the network.

## Status of This Memo

Table of Contents

## 1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload the path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440].

This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCE-based Central Controller (PCECC) architecture.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment Routing (SR) technology leverages the source routing and tunneling paradigms. A source node can choose a path without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [RFC8402] provides an introduction to SR architecture. The corresponding IS-IS and OSPF extensions are specified in [RFC8667] and [RFC8665] , respectively. It relies on a series of forwarding instructions being placed in the header of a packet. The list of segments forming the path is called the Segment List and is encoded in the packet header. Segment Routing can be applied to the IPv6 architecture with the Segment Routing Header (SRH) [RFC8754]. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment. The segment routing architecture supports operations that can be used to steer packet flows in a network, thus providing a form of traffic engineering. [RFC8664] and [I-D.ietf-pce-segment-routing-ipv6] specify the SR specific PCEP extensions.

PCECC may further use PCEP for SR SID (Segment Identifier) distribution on the SR nodes with some benefits. [I-D.zhao-pce-pcep-extension-pce-controller-sr] specifies the procedures and PCEP extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR-MPLS SID distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network. This document extends this to include SRv6 SID distribution as well.

### 1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 2. Terminology

Terminologies used in this document is the same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

## 3. PCECC SRv6

[RFC8664] specifies extensions to PCEP that allow a stateful PCE to compute, update, or initiate SR-TE paths for MPLS dataplane. An ingress node of an SR-TE path appends all outgoing packets with a list of MPLS labels (SIDs). This is encoded in SR-ERO subobject, capable of carrying a label (SID) as well as the identity of the node/adjacency label (SID). [I-D.ietf-pce-segment-routing-ipv6] extends the procedure to include support for SRv6 paths.

As per [RFC8754], an SRv6 Segment is a 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment". Further details are in an illustration provided in [I-D.ietf-spring-srv6-network-programming]. The SR is applied to IPV6 data plane using SRH. An SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool, or a PCE and provisioned on the ingress node. [I-D.ietf-pce-segment-routing-ipv6] specify the SRv6-ERO subobject capable of carrying an SRv6 SID as well as the identity of the node/adjacency represented by the SID.

As per [RFC8283], PCECC can allocate and provision the node/prefix/adjacency label (SID) via PCEP. As per [I-D.ietf-teas-pcecc-use-cases] this is also applicable to SRv6 SIDs.

The rest of the processing is similar to existing stateful PCE for SRv6 [I-D.ietf-pce-segment-routing-ipv6].

## 4. PCEP Requirements

Following key requirements for PCECC-SRv6 should be considered when designing the PCECC-based solution:

- A PCEP speaker supporting this draft needs to have the capability to advertise its PCECC-SRv6 capability to its peers.
- PCEP procedures need to allow for PCC-based SRv6 SID allocations.
- PCEP procedures need means to update (or clean up) the SRv6 SID to the PCC.
- PCEP procedures need to provide a mean to synchronize the SRv6 SID allocations between the PCE to the PCC in the PCEP messages.

## 5. Procedures for Using the PCE as a Central Controller (PCECC) in SRv6

### 5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a Central Controller (PCECC) reuses the existing active stateful PCE mechanism as much as possible to control the LSPs.

### 5.2. New Functions

This document uses the same PCEP messages and its extensions which are described in [I-D.ietf-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr] for PCECC-SRv6 as well.

The PCEP messages PCRpt, PCInitiate, PCUpd are used to send LSP Reports, LSP setup, and LSP update respectively. The extended PCInitiate message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is used to download or clean up central controller's instructions (CCIs) (SRv6 SID in the scope of this document). The extended PCRpt message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is also used to report the CCIs (SRv6 SIDs) from PCC to PCE.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify an object called CCI for the encoding of the central controller's instructions. [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object-type for segment routing. This document further defines a new CCI object-type for SRv6.

### 5.3. PCECC Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions. A PCEP Speaker includes the "PCECC Capability" sub-TLV, described in [I-D.ietf-pce-pcep-extension-for-pce-controller].

A new S-bit is added in the PCECC-CAPABILITY sub-TLV to indicate support for PCECC-SR in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. This document adds another I-bit to indicate support for SR in IPv6. A PCC MUST set the I-bit in the PCECC-CAPABILITY sub-TLV and include the SRv6-PCE-CAPABILITY sub-TLV ([I-D.ietf-pce-segment-routing-ipv6]) in the OPEN Object (inside the PATH-SETUP-TYPE-CAPABILITY TLV) to support the PCECC SRv6 extensions defined in this document. If I-bit is set in PCECC-CAPABILITY sub-TLV and the SRv6-PCE-CAPABILITY sub-TLV is not advertised in the OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD4 (SRv6 capability was not advertised) and terminate the session.

The rest of the processing is as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

### 5.4. PCEP session IP address and TED Router ID

As described in [I-D.zhao-pce-pcep-extension-pce-controller-sr], it is important to link the session IP address with the Router ID in TED for successful PCECC-SRv6 operations.

### 5.5. SRv6 Path Operations

[RFC8664] specify the PCEP extension to allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks. [I-D.ietf-pce-segment-routing-ipv6] extends it to support SRv6.

The Path Setup Type for SRv6 (PST=TBD) is used on the PCEP session with the Ingress as per [I-D.ietf-pce-segment-routing-ipv6].

#### 5.5.1. PCECC Segment Routing in IPv6 (SRv6)

Segment Routing (SR) as described in [RFC8402] depends on "segments" that are advertised by Interior Gateway Protocols (IGPs). The SR-node allocates and advertises the SID (node, adj, etc) and floods them via the IGP. This document proposes a new mechanism where PCE allocates the SRv6 SID centrally and uses PCEP to advertise them. In some deployments, PCE (and PCEP) are better suited than IGP because of the centralized nature of PCE and direct TCP based PCEP sessions to the node.

#### 5.5.1.1. PCECC SRv6 Node/Prefix SID allocation

Each node (PCC) is allocated a node SRv6 SID by the PCECC. The PCECC sends the PCInitiate message to update the SRv6 SID table of each node. The TE router ID is determined from the TED or from "IPv4/IPv6 Router-ID" Sub-TLV [I-D.dhodylee-pce-pcep-ls], in the OPEN Object.

On receiving the SRv6 node SID allocation, each node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly. The PCInitiate message uses the FEC object [I-D.zhao-pce-pcep-extension-pce-controller-sr].

On receiving the SRv6 node SID allocation:

For the local SID, the node (PCC) needs to update SID with associated function (END function in this case) in "My Local SID Table" ([I-D.ietf-spring-srv6-network-programming]).

For the non-local SID, the node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly.

The forwarding behavior and the end result is similar to IGP based "Node-SID" in SRv6. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node as per [RFC8402].

PCE relies on the Node/Prefix SRv6 SID clean up using the same PCInitiate message as per [RFC8281].

#### 5.5.1.2. PCECC SRv6 Adjacency SID allocation

For PCECC-SRv6, apart from node-SID, Adj-SID is used where each adjacency is allocated an Adj-SID by the PCECC. The PCECC sends PCInitiate message to update the SRv6 SID entry for each adjacency to the corresponding nodes in the domain. Each node (PCC) download the SRv6 SID instructions accordingly. Similar to SRv6 Node/Prefix Label allocation, the PCInitiate message in this case uses the FEC object.

The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SRv6 as per [RFC8402].

The handling of adjacencies on the LAN subnetworks is specified in [RFC8402]. PCECC MUST assign Adj-SID for every pair of routers in the LAN. The rest of the protocol mechanism remains the same.

PCE relies on the Adj label clean up using the same PCInitiate message as per [RFC8281].

#### 5.5.1.3. Redundant PCEs

[I-D.litkowski-pce-state-sync] describes the synchronization mechanism between the stateful PCEs. The SRv6 SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC-SRv6 state synchronization. Note that the SRv6 SIDs are independent of the SRv6 paths, and remains intact till any topology change. The redundant PCEs MUST have a common view of all SRv6 SIDs allocated in the domain.

#### 5.5.1.4. Re-Delegation and Clean up

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the action needed for CCIs for the static LSPs on a terminated session. Same holds true for the CCI for SRv6 SID as well.

#### 5.5.1.5. Synchronization of SRv6 SID Allocations

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the synchronization of Central Controller's Instructions (CCI) via the LSP state synchronization as described in [RFC8231] and [RFC8232]. Same procedures are applied for the CCI for the SRv6 SIDs as well.

# 6. PCEP Messages

The PCEP messages are as per [I-D.zhao-pce-pcep-extension-pce-controller-sr].

# 7. PCEP Objects

## 7.1. OPEN Object

### 7.1.1. PCECC Capability sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defined the PCECC-CAPABILITY sub-TLV.

A new I-bit is defined in PCECC-CAPABILITY sub-TLV for PCECC-SRv6:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Type=TBD            |           Length=4            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Flags                         |I|S|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

[Editor's Note - The above figure is included for ease of the reader but should be removed before publication.]

I (PCECC-SRv6-CAPABILITY - 1 bit - TBD1): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable of PCECC-SRv6 capability and the PCE allocates the Node and Adj SRv6 SID on this session.

## 7.2. SRv6 Path Setup

The PATH-SETUP-TYPE TLV is defined in [RFC8408]. A PST value of TBD is used when Path is setup via SRv6 mode as per [I-D.ietf-pce-segment-routing-ipv6]. The procedure for SRv6 path setup as specified in [I-D.ietf-pce-segment-routing-ipv6] remians unchanged.

## 7.3. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the cntroller instructions is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]. This document defines another object-type for SRv6 purpose.

CCI Object-Type is TBD3 for SRv6 as below -

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             CC-ID                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      MT-ID    |   Algorithm   |    Flags      |B|P|G|C|N|E|V|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Reserved            |   SRv6 Endpoint Function      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                      SRv6 Identifier                          |
|                         (128-bit)                             |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                        Optional TLV                         //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The field CC-ID is as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. The field MT-ID, Algorithm, Flags are defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

Reserved: MUST be set to 0 while sending and ignored on receipt.

SRv6 Endpoint Function: 16-bit field representing supported functions associated with SRv6 SIDs.

SRv6 Identifier: 128-bit IPv6 addresses representing SRv6 segment.

[Editor's Note - It might be useful to separate the LOC:FUNC part in the SRv6 SID (future study)]

### 7.4. FEC Object

The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object (and various Object-Types) are described in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. SRv6 Node SID MUST includes the FEC Object-Type 2 for IPv6 Node. SRv6 Adjacency SID MUST include the FEC Object-Type=4 for IPv6 adjacency. Further FEC object types could be added in future extensions.

## 8. Security Considerations

The security considerations described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

## 9. Manageability Considerations

### 9.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/ disable PCECC SRv6 capability as a global configuration.

### 9.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC SRv6 capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC SRv6 capability.

### 9.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

### 9.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

### 9.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

### 9.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCInitiate/PCUpd messages (as per [RFC8231]) sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

## 10. IANA Considerations

### 10.1. PCECC-CAPABILITY sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY sub-TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field registry, as follows:

| Bit | Description | Reference |
|---|---|---|
| TBD1 | SRv6 | This document |

### 10.2. PCEP Object

IANA is requested to allocate a new code-point for the new CCI object-type in "PCEP Objects" sub-registry as follows:

| Object-Class Value | Name | Object-Type | Reference |
|---|---|---|---|
| TBD | CCI | | |
| | | TBD3: SRv6 | This document |

### 10.3. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

| Error-Type | Meaning | |
|---|---|---|
| 19 | Invalid operation. | |
| | Error-value = TBD4 : | SRv6 capability was not advertised |

## 11. Acknowledgments

## 12. References

### 12.1. 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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[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>.

[I-D.ietf-pce-segment-routing-ipv6] Li, C., Negi, M., Koldychev, M., Kaladharan, P., and Y. Zhu, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-ietf-pce-segment-routing-ipv6-06 (work in progress), July 2020.

[I-D.ietf-pce-pcep-extension-for-pce-controller] Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-ietf-pce-pcep-extension-for-pce-controller-07 (work in progress), September 2020.

[I-D.zhao-pce-pcep-extension-pce-controller-sr]
Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-07 (work in progress), July 2020.

## 12.2. Informative References

[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>.

[RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <https://www.rfc-editor.org/info/rfc7025>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <https://www.rfc-editor.org/info/rfc7399>.

[RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <https://www.rfc-editor.org/info/rfc7420>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <https://www.rfc-editor.org/info/rfc7491>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <https://www.rfc-editor.org/info/rfc7525>.

[RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <https://www.rfc-editor.org/info/rfc8232>.

[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>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[I-D.ietf-teas-pcecc-use-cases] Li, Z., Khasanov, B., Dhody, D., Zhao, Q., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-06 (work in progress), September 2020.

[I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-14 (work in progress), July 2020.

[I-D.litkowski-pce-state-sync]
Litkowski, S., Sivabalan, S., Li, C., and H. Zheng, "Inter Stateful Path Computation Element (PCE) Communication Procedures.", draft-litkowski-pce-state-sync-08 (work in progress), July 2020.

[I-D.dhodylee-pce-pcep-ls]
Dhody, D., Peng, S., Lee, Y., Ceccarelli, D., and A. Wang, "PCEP extensions for Distribution of Link-State and TE Information", draft-dhodylee-pce-pcep-ls-17 (work in progress), July 2020.

[I-D.ietf-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-ietf-spring-srv6-network-programming-24 (work in progress), October 2020.

## Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

## Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: pengshuping@huawei.com

Xuesong Geng
Huawei Technologies
China

EMail: gengxuesong@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore, Karnataka 560102
India

EMail: mahend.ietf@gmail.com

PCE Working Group D. Dhody, Ed.
Internet-Draft Huawei Technologies
Intended status: Standards Track S. Sivabalan
Expires: April 22, 2019 Cisco Systems, Inc.
October 19, 2018

# A YANG Data Model for Segment Routing in IPv6 (SRv6) support in Path Computation Element Communications Protocol (PCEP)

draft-dhody-pce-pcep-srv6-yang-00

## Abstract

This document augments a YANG data model for the management of Path Computation Element communications Protocol (PCEP) for communications between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs in support for Segment Routing in IPv6. The data model includes configuration data and state data (status information and counters for the collection of statistics).

## Status of This Memo



## Copyright Notice

Table of Contents



## 1. Introduction

The Path Computation Element (PCE) defined in [RFC4655] is an entity that is capable of computing a network path or route based on a network graph, and applying computational constraints. A Path Computation Client (PCC) may make requests to a PCE for paths to be computed.

PCEP is the communication protocol between a PCC and PCE and is defined in [RFC5440]. PCEP interactions include path computation requests and path computation replies as well as notifications of specific states related to the use of a PCE in the context of Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering (TE). [RFC8231] specifies extensions to PCEP to enable stateful control of MPLS TE LSPs.

[I-D.negi-pce-segment-routing-ipv6] extends [I-D.ietf-pce-segment-routing] to support SR for IPv6 data plane.

[I-D.ietf-pce-pcep-yang] defines a YANG [RFC7950] data model for the management of PCEP speakers. This document contains a specification of the PCEP-SRv6 YANG module, "ietf-pcep-srv6" which provides the PCEP-SRv6 [I-D.negi-pce-segment-routing-ipv6] data model.

The PCEP operational state is included in the same tree as the PCEP configuration consistent with Network Management Datastore Architecture [RFC8342]. The origin of the data is indicated as per the origin metadata annotation.

## 2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 3. Terminology and Notation

This document also uses the following terms defined in [RFC7420]:

- PCEP entity: a local PCEP speaker.
- PCEP peer: to refer to a remote PCEP speaker.
- PCEP speaker: where it is not necessary to distinguish between local and remote.

Further, this document also uses the following terms defined in [RFC8231] :

- Stateful PCE, Passive Stateful PCE, Active Stateful PCE
- Delegation, Revocation, Redelegation
- LSP State Report, Path Computation Report message (PCRpt).
- LSP State Update, Path Computation Update message (PCUpd).

[RFC8281] :

- PCE-initiated LSP, Path Computation LSP Initiate Message (PCInitiate).

[RFC8408] :

- Path Setup Type (PST).

[I-D.ietf-pce-segment-routing] :

- Segment Routing (SR).

[I-D.negi-pce-segment-routing-ipv6] :

- o Segment Routing in IPv6 (SRv6).

### 3.1. Tree Diagrams

A simplified graphical representation of the data model is used in this document. The meaning of the symbols in these diagrams is defined in [RFC8340].

### 3.2. Prefixes in Data Node Names

In this document, names of data nodes and other data model objects are often used without a prefix, as long as it is clear from the context in which YANG module each name is defined. Otherwise, names are prefixed using the standard prefix associated with the corresponding YANG module, as shown in Table 1.

| Prefix | YANG module | Reference |
|---|---|---|
| te-types | ietf-te-types | [I-D.ietf-teas-yang-te] |
| p | ietf-pcep | [I-D.ietf-pce-pcep-yang] |
| st | ietf-srv6-types | [I-D.raza-spring-srv6-yang] |

Table 1: Prefixes and corresponding YANG modules

## 4. The Design of PCEP-SRv6 Data Model

### 4.1. The Overview of PCEP SRv6 Data Model

The PCEP-SRv6 YANG module defined in this document has all the common building blocks for the PCEP-SRv6 extention.

```
   module: ietf-pcep-srv6
     augment /p:pcep/p:entity/p:capability:
       +--rw srv6 {srv6}?
          +--rw enabled?    boolean
          +--rw max-sl?     uint8
          +--rw sl-limit?   boolean
     augment /p:pcep/p:entity/p:peers/p:peer/p:capability:
       +--rw srv6 {srv6}?
          +--rw enabled?    boolean
          +--rw max-sl?     uint8
          +--rw sl-limit?   boolean
     augment /p:pcep/p:entity/p:lsp-db/p:lsp:
       +--ro srv6 {srv6}?
          +--ro segment-list
             +--ro segment* [index]
                +--ro index        uint32
                +--ro sid-value?   st:srv6-sid
     groupings:
     segment-list
         +---- segment-list
            +---- segment* [index]
               +---- index?       uint32
               +---- sid-value?   st:srv6-sid

     segment-properties
         +---- index?       uint32
         +---- sid-value?   st:srv6-sid

     srv6
         +---- srv6 {srv6}?
            +---- enabled?    boolean
            +---- max-sl?     uint8
            +---- sl-limit?   boolean
```

5. PCEP-SRv6 YANG Modules

5.1. ietf-pcep-srv6 module

RFC Ed.: In this section, replace all occurrences of 'XXXX' with the actual RFC number and all occurrences of the revision date below with the date of RFC publication (and remove this note).

```
   <CODE BEGINS> file "ietf-pcep-srv6@2018-10-19.yang"
   module ietf-pcep-srv6 {

       yang-version 1.1;
```

```
       namespace "urn:ietf:params:xml:ns:yang:ietf-pcep-srv6";
       prefix ps;


       import ietf-srv6-types {
           prefix "st";
           reference "RFC XXXX";
       }

       import ietf-te-types {
          prefix "te-types";
          reference "RFC XXXX";
       }

       import ietf-pcep {
          prefix "p";
          reference "RFC XXXX";
       }


       organization
           "IETF PCE (Path Computation Element) Working Group";

       contact
           "WG Web:   <http://tools.ietf.org/wg/pce/>
            WG List:  <mailto:pce@ietf.org>
            Editor:   Dhruv Dhody
                      <mailto:dhruv.ietf@gmail.com>";
       description
           "The YANG module augments the PCEP yang operational
           model with SRv6";


       revision 2018-10-19 {
           description "Initial revision.";
           reference
               "RFC XXXX:  A YANG Data Model for Path Computation
                           Element Communications Protocol
                           (PCEP) - Segement Routing in IPv6
                           (SRv6)";
       }

       /* Identity */
       identity path-setup-srv6 {
           base te-types:path-signaling-type;
           description
             "SRv6 path setup type";
       }
```

```
    /* Features */
    feature srv6 {
        description
            "Support Segment Routing in IPv6 (SRv6) for PCE.";
    }

    /* Groupings */
    grouping srv6 {
        description
            "SRv6";
        container srv6 {
            if-feature srv6;
            description
                "If SRv6 is supported";
            leaf enabled{
                type boolean;
                description
                    "Enabled or Disabled";
            }
            leaf max-sl {
                type uint8;
                description
                    "Max value of the segment left field in SRH";
            }
            leaf sl-limit{
                type boolean;
                default false;
                description
                    "True indicates no limit on SL, the
                    leaf max-sl is ignored";
            }
        }
    }

    grouping segment-list {
        description
            "Segment list grouping";
        container segment-list {
            description
                "Segments for given segment list";

            list segment {
                key "index";
                description "Configure Segment/hop at the index";
                uses segment-properties;
            }
        }
    }
```

```
     grouping segment-properties {
         description "Segment properties grouping";
         leaf index {
             type uint32;
             description "Segment index";
         }
         leaf sid-value {
             type st:srv6-sid;
             description "SRv6 SID value";
         }
     }


     /*
      * Augment modules to add SRv6
      */
     augment "/p:pcep/p:entity/p:capability"{
         description
             "Augmenting SRv6";
         uses srv6;
     }
     augment "/p:pcep/p:entity/p:peers/p:peer/p:capability"{
         description
             "Augmenting SRv6";
         uses srv6;
     }
     augment "/p:pcep/p:entity/p:lsp-db/p:lsp"{
         description
             "Augmenting SRv6";
         container srv6 {
             when "/p:pcep/p:entity/p:lsp-db/p:lsp/p:pst
                   = 'path-setup-srv6'" {
                 description
                     "For SRv6 path";
             }
             if-feature srv6;
             uses segment-list;
             description
                 "SRv6";
         }
     }

 }//module

 <CODE ENDS>
```

## 6. Security Considerations

The YANG module defined in this document is designed to be accessed via network management protocol such as NETCONF [RFC6241] or RESTCONF [RFC8040]. The lowest NETCONF layer is the secure transport layer and the mandatory-to-implement secure transport is SSH [RFC6242]. The lowest RESTCONF layer is HTTPS, and the mandatory-to-implement secure transport is TLS [RFC8446]

The NETCONF access control model [RFC8341] provides the means to restrict access for particular NETCONF or RESTCONF users to a pre-configured subset of all available NETCONF or RESTCONF protocol operations and content.

There are a number of data nodes defined in the YANG module which are writable/creatable/deletable (i.e., config true, which is the default). These data nodes may be considered sensitive or vulnerable in some network environments. Write operations (e.g., <edit-config>) to these data nodes without proper protection can have a negative effect on network operations. These are the subtrees and data nodes and their sensitivity/vulnerability:

/p:pcep/p:entity/p:capability/ps:srv6 - configure local SRv6 capability and parameters.

/p:pcep/p:entity/p:peers/p:peer/p:capability/ps:srv6 - configure peer's SRv6 capability and parameters.

Unauthorized access to above list can adversely affect the PCEP session between the local entity and the peers. This may lead to inability to compute new paths, stateful operations on the delegated as well as PCE-initiated LSPs.

Some of the readable data nodes in this YANG module may be considered sensitive or vulnerable in some network environments. It is thus important to control read access (e.g., via get, get-config, or notification) to these data nodes. These are the subtrees and data nodes and their sensitivity/vulnerability:

/p:pcep/p:entity/p:lsp-db/p:lsp/ps:srv6 - The SRv6 SID in the network. Unauthorized access to this could provide the all path and network usage information.

## 7. IANA Considerations

This document registers a URI in the "IETF XML Registry" [RFC3688]. Following the format in RFC 3688, the following registration has been made.

URI: urn:ietf:params:xml:ns:yang:ietf-pcep-srv6

Registrant Contact: The PCE WG of the IETF.

XML: N/A; the requested URI is an XML namespace.

This document registers a YANG module in the "YANG Module Names" registry [RFC6020].

```
Name:         ietf-pcep
Namespace:    urn:ietf:params:xml:ns:yang:ietf-pcep-srv6
Prefix:       ps
Reference:    This I-D
```

## 8. Acknowledgements

## 9. References

### 9.1. 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>.

[RFC3688] Mealling, M., "The IETF XML Registry", BCP 81, RFC 3688, DOI 10.17487/RFC3688, January 2004, <https://www.rfc-editor.org/info/rfc3688>.

[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>.

[RFC6020] Bjorklund, M., Ed., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)", RFC 6020, DOI 10.17487/RFC6020, October 2010, <https://www.rfc-editor.org/info/rfc6020>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC6242] Wasserman, M., "Using the NETCONF Protocol over Secure Shell (SSH)", RFC 6242, DOI 10.17487/RFC6242, June 2011, <https://www.rfc-editor.org/info/rfc6242>.

[RFC7950] Bjorklund, M., Ed., "The YANG 1.1 Data Modeling Language", RFC 7950, DOI 10.17487/RFC7950, August 2016, <https://www.rfc-editor.org/info/rfc7950>.

[RFC8040] Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", RFC 8040, DOI 10.17487/RFC8040, January 2017, <https://www.rfc-editor.org/info/rfc8040>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[RFC8340] Bjorklund, M. and L. Berger, Ed., "YANG Tree Diagrams", BCP 215, RFC 8340, DOI 10.17487/RFC8340, March 2018, <https://www.rfc-editor.org/info/rfc8340>.

[RFC8341] Bierman, A. and M. Bjorklund, "Network Configuration Access Control Model", STD 91, RFC 8341, DOI 10.17487/RFC8341, March 2018, <https://www.rfc-editor.org/info/rfc8341>.

[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>.

[RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018, <https://www.rfc-editor.org/info/rfc8446>.

[I-D.ietf-teas-yang-te] Saad, T., Gandhi, R., Liu, X., Beeram, V., Shah, H., and I. Bryskin, "A YANG Data Model for Traffic Engineering Tunnels and Interfaces", draft-ietf-teas-yang-te-16 (work in progress), July 2018.

[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-14 (work in progress), October 2018.

[I-D.raza-spring-srv6-yang]
Raza, K., Rajamanickam, J., Liu, X., Hu, Z., Hussain, I., Shah, H., daniel.voyer@bell.ca, d., Elmalky, H., Matsushima, S., Horiba, K., and A. Abdelsalam, "YANG Data Model for SRv6 Base and Static", draft-raza-spring-srv6-yang-01 (work in progress), March 2018.

[I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-09 (work in progress), October 2018.

[I-D.negi-pce-segment-routing-ipv6]
Negi, M., Kaladharan, P., Dhody, D., and S. Sivabalan, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-negi-pce-segment-routing-ipv6-02 (work in progress), June 2018.

## 9.2. Informative References

[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>.

[RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <https://www.rfc-editor.org/info/rfc7420>.

[RFC8342] Bjorklund, M., Schoenwaelder, J., Shafer, P., Watsen, K., and R. Wilton, "Network Management Datastore Architecture (NMDA)", RFC 8342, DOI 10.17487/RFC8342, March 2018, <https://www.rfc-editor.org/info/rfc8342>.

## Authors' Addresses

Dhruv Dhody (editor)
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka  560066
India

EMail: dhruv.ietf@gmail.com

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario  K2K 3E8
Canada

EMail: msiva@cisco.com

PCE Working Group
Internet Draft
Intended status: Standard Track
Expires: April 23, 2019

Y. Lee (Editor)
H. Zheng
Huawei

R. Casellas
R. Vilalta
CTTC

D. Ceccarelli
F. Lazzeri
Ericsson

October 22, 2018

# PCEP Extension for Flexible Grid Networks

draft-lee-pce-flexible-grid-03

## Abstract

This document provides the Path Computation Element Communication Protocol (PCEP) extensions for the support of Routing and Spectrum Assignment (RSA) in Flexible Grid networks.

## Status of this Memo

This Internet-Draft will expire on December 22, 2018.



Table of Contents

## 1. Terminology

This document uses the terminology defined in [RFC4655], [RFC5440] and [RFC7698].

## 2. 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 [RFC2119].

## 3. Introduction

[RFC4655] defines a PCE based path computation architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is said to be any network component that makes such a request and may be, for instance, an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communications Protocol (PCEP) is the communication protocol used between a PCC and a PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

[PCEP-WSON] provides the PCEP extensions for the support of Routing and Wavelength Assignment (RWA) in Wavelength Switched Optical Networks (WSON) based on the requirements specified in [RFC6163] and [RFC7449].

[RFC7698] provides Framework and Requirements for GMPLS-Based Control of Flexi-Grid Dense Wavelength Division Multiplexing (DWDM)

Networks. To allow efficient allocation of optical spectral bandwidth for systems that have high bit-rates, the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) has extended its Recommendations G.694.1 and G.872 to include a new Dense Wavelength Division Multiplexing (DWDM) grid by defining a set of nominal central frequencies, channel spacings, and the concept of the "frequency slot". In such an environment, a data-plane connection is switched based on allocated, variable-sized frequency ranges within the optical spectrum, creating what is known as a flexible grid (flexi-grid).

This document provides PCEP extensions to support Routing and Spectrum Assignment (RSA) in in Spectrum Switched Optical Networks (SSON)[RFC7698].

Figure 2 shows one typical PCE based implementation, which is referred to as the Combined Routing and Spectrum Assignment (R&SA) [RFC7698]. With this architecture, the two processes of routing and spectrum assignment are accessed via a single PCE. This architecture is the base architecture from which the PCEP extensions are going to be specified in this document.

```
                       +-----------------------------+
         +-----+       |     +-------+      +--+     |
         |     |       |     |Routing|      |SA|     |
         | PCC |<----->|     +-------+      +--+     |
         |     |       |                             |
         +-----+       |             PCE             |
                       +-----------------------------+
```

Figure 1 Combined Process (R&SA) architecture

## 4. Spectrum Assignment (SA) Object

Spectrum allocation can be performed by the PCE by different means:

(a) By means of Explicit Label Control (ELC) where the PCE allocates which label to use for each interface/node along the path.

(b) By means of a Label Set where the PCE provides a range of potential frequency slots to allocate by each node along the path. This document aligns with GMPLS extensions for PCEP [PCEP-GMPLS] for generic property such as label, label-set and label assignment noting that frequency is a type of label. Frequency restrictions and constraints are also formulated in terms of labels per [RFC7579].

Option (b) allows distributed spectrum allocation (performed during signaling) to complete spectrum assignment.

Additionally, given a range of potential spectrums to allocate, the request SHOULD convey the heuristic / mechanism to the allocation.

The format of a PCReq message after incorporating the Spectrum Assignment (SA) object is as follows:

```
<PCReq Message> ::= <Common Header>

                       [<svec-list>]

                       <request-list>

   Where:

      <request-list>::=<request>[<request-list>]

      <request>::= <RP>

                   <GENERALIZED ENDPOINTS>

                   [ <SA> ]

                   [other optional objects...]
```

If the SA object is present in the request, it MUST be encoded after the ENDPOINTS object.

The format of the Spectrum Assignment (SA) object body is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Reserved             |             Flags           |M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Frequency-Slot Selection TLV                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Frequency-Slot Restriction Constraint TLV             |
```

```
   .                                                               .
   .                                                               .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //                      Optional TLVs                          //
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2 SA Object

o Reserved (16 bits)

o Flags (16 bits)

The following new flags SHOULD be set

. M (Mode - 1 bit): M bit is used to indicate the mode of spectrum assignment. When M bit is set to 1, this indicates that the spectrum assigned by the PCE must be explicit. That is, the selected way to convey the allocated spectrum is by means of Explicit Label Control (ELC) [RFC4003] for each hop of a computed LSP. Otherwise, the spectrum assigned by the PCE needs not be explicit (i.e., it can be suggested in the form of label set objects in the corresponding response, to allow distributed SA. In such case, the PCE MUST return a Label Set Field as described in Section 2.6 of [RFC7579] in the response. See Section 5 of this document for the encoding discussion of a Label Set Field in a PCRep message.

## 4.1. Frequency-Slot Selection TLV

The Frequency-Slot Selection TLV is used to indicate the frequency-slot selection constraint in regard to the order of frequency-slot assignment to be returned by the PCE. This TLV is only applied when M bit is set in the SA Object specified in Section 3.1. This TLV MUST NOT be used when the M bit is cleared.

The Frequency-Slot Selection sub-TLV value field is defined as:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|S|  FSA Method |                   Reserved                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

S (Symmetry, 1 bit): This flag is only meaningful when the request is for a bidirectional LSP (see [RFC5440]).

0 denotes requiring the same frequency-slot in both directions; 1 denotes that different spectrums on both directions are allowed.

Frequency-Slot Assignment (FSA) Method (7 bits):

0: unspecified (any); This does not constrain the SA method used by a PCC  This value is implied when the Frequency-Slot Selection sub-TLV is absent.

1: First-Fit. All the feasible frequency slots are numbered (based on "n" parameter), and this SA method chooses the available frequency-slot with the lowest index (of "n" parameter).

2: Random. This SA method chooses an feasible frequency-slot ("n" paramerer) randomly.

3-127: Unassigned.

The processing rules for this TLV are as follows:

If a PCE does not support the attribute(s), its behavior is specified below:

- S bit not supported: a PathErr MUST be generated with the Error Code "Routing Problem" (24) with error sub-code "Unsupported Frequency slot Selection Symmetry value" (TDB).

- FSA method not supported: a PathErr MUST be generated with the Error Code "Routing Problem" (24) with error sub-code

"Unsupported Frequency Slot Assignment value" (TDB).

A Frequency Slot Selection TLV can be constructed by a node and added to an ERO Hop Attributes subobject in order to be processed by downstream nodes (transit and egress).  As defined in [RFC7570], the R bit reflects the LSP_REQUIRED_ATTRIBUTE and LSP_ATTRIBUTE semantic defined in [RFC5420], and it SHOULD be set accordingly.

Once a node properly parses the Spectrum Selection sub-TLV received in an ERO Hop Attributes subobject, the node use the indicated spectrum assignment method (at that hop) for the LSP. In addition, the node SHOULD report compliance by adding an RRO Hop Attributes subobject with the WSON Processing Hop Attribute TLV (and its sub-TLVs) that indicate the utilized method. Frequency-Slot Selection TLVs carried in an RRO Hop Attributes subobject are subject to [RFC7570] and standard RRO processing; see [RFC3209].

## 4.2. Frequency-slot Restriction Constraint TLV

For any request that contains a Frequency-slot assignment, the requester (PCC) MUST be able to specify a restriction on the frequency-slots to be used. This restriction is to be interpreted by the PCE as a constraint on the tuning ability of the origination laser transmitter or on any other maintenance related constraints.

The format of the Frequency-Slot Restriction Constraint TLV is as follows:

```
<Frequency-lot Restriction Constraint> ::=

               <Action> <Count> <Reserved>

               (<Link Identifiers> <Freq-slot Restriction>)...
```

Where

```
<Link Identifiers> ::= <Link Identifier> [<Link Identifiers>]
```

See Section 4.3.1 in [PCEP-WSON] for the encoding of the Link Identifiers Field.

The Frequency slot Restriction Constraint TLV type is TBD. This TLV MAY appear more than once to be able to specify multiple restrictions.

The TLV data is defined as follows:

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Action            |   Count       |          Reserved           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                     Link Identifiers                          |
 |                          . . .                                |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Frequency Slot Restriction Field               |
 //                        . . . .                              //
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3 spectrum Restriction Constraint TLV Encoding

o  Action: 8 bits

  . 0 - Inclusive List indicates that one or more link identifiers are included in the Link Set. Each identifies a separate link that is part of the set.

  . 1 - Inclusive Range indicates that the Link Set defines a range of links.  It contains two link identifiers. The first identifier indicates the start of the range (inclusive). The second identifier indicates the end of the range (inclusive). All links with numeric values between the bounds are considered to be part of the set. A value of zero in either position indicates that there is no bound on the corresponding portion of the range. Note that the Action field can be set to 0 when unnumbered link identifier is used.

o  Count: The number of the link identifiers (8 bits)

Note that a PCC MAY add a spectrum restriction that applies to all links by setting the Count field to zero and specifying just a set of spectrums.

Note that all link identifiers in the same list must be of the same type.

o  Reserved: Reserved for future use (16 bits)

o  Link Identifiers: Identifies each link ID for which restriction is applied. The length is dependent on the link format and the Count field. See Section 4.3.1 in [PCEP-WSON] for Link Identifier encoding and Section 3.3.1 for the Spectrum Restriction Field encoding, respectively.

### 4.2.1. Frequency-Slot Restriction Field

The Frequency-Slot Restriction Field of the Frequency slot restriction TLV is encoded as defined in https://tools.ietf.org/html/draft-ietf-ccamp-flexible-grid-ospf-ext-09#section-4.1.1.

## 5. Encoding of a RSA Path Reply

This section provides the encoding of a RSA Path Reply for frequency slot allocation as discussed in Section 4. Spectrum Allocation TLV

The Spectrum Allocation TLV type is TBD, recommended value is TBD. The TLV data is defined as follows:

```
 0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |               Type            |           Length            |M|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                          Link Identifier                      |
 |                               . . .                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
|                    Allocated Spectrum(s)                      |
//                         . . . .                             //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4 Spectrum Allocation TLV Encoding

- o Type (16 bits): The type of the TLV.

- o Length (15 bits): The length of the TLV including the Type and Length fields.

- o M (Mode): 1 bit

  - 0 indicates the allocation is under Explicit Label Control.

  - 1 indicates the allocation is expressed in Label Sets.

Note that all link identifiers in the same list must be of the same type.

o Link Identifier (variable): Identifies the interface to which assignment spectrum(s) is applied. See Section 3.3 for Link Identifier encoding.

o Allocated Spectrum(s) (variable): Indicates the allocated spectrum(s) to the link identifier. See Section 3.3.1 for encoding details.

This TLV is encoded as an attributes TLV, per [RFC5420], which is carried in the ERO LSP Attribute Subobjects per [RFC7570]. The type value of the Spectrum Restriction Constraint TLV is TBD by IANA.

## 5.1. Error Indicator

To indicate errors associated with the RSA request, a new Error Type (TDB) and subsequent error-values are defined as follows for inclusion in the PCEP-ERROR Object:

A new Error-Type (TDB) and subsequent error-values are defined as follows:

- Error-Type=TBD; Error-value=1: if a PCE receives a RSA request and the PCE is not capable of processing the request due to insufficient memory, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=TDB) and an Error-value(Error-value=1). The PCE stops processing the request. The corresponding RSA request MUST be cancelled at the PCC.

- Error-Type=TBD; Error-value=2: if a PCE receives a RSA request and the PCE is not capable of RSA computation, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=TDB) and an Error-value (Error-value=2). The PCE stops processing the request. The corresponding RSA computation MUST be cancelled at the PCC.

## 5.2. NO-PATH Indicator

To communicate the reason(s) for not being able to find RSA for the path request, the NO-PATH object can be used in the corresponding response. The format of the NO-PATH object body is defined in [RFC5440]. The object may contain a NO-PATH-VECTOR TLV to provide additional information about why a path computation has failed.

One new bit flag is defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object.

- Bit TDB: When set, the PCE indicates no feasible route was found that meets all the constraints (e.g., spectrum restriction, etc.) associated with RSA.

# 6. Manageability Considerations

Manageability of SSON Routing and Spectrum Assignment (RSA) with PCE must address the following considerations:

## 6.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- The ability to send a Flexi-Grid RSA request.

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- The support for Flexi-Grid RSA .
- A set of Flexi-Grid RSA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

## 6.2. Information and Data Models

Extensions to the PCEP YANG module may include to cover the Flexi-Grid RSA information introduced in this document. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [RFC5440].

## 6.3. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [RFC5440]

## 6.4. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise Flexi-Grid RSA path computation capabilities to PCCs.

This draft has requirements on other protocols (ERO objects, etc. which are under TEAS or CCAMP.)

## 6.5. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

# 7. Security Considerations

This document has no requirement for a change to the security models within PCEP. However, the additional information distributed in order to address the RSA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

# 8. IANA Considerations

IANA maintains a registry of PCEP parameters. IANA has made allocations from the sub-registries as described in the following sections.

## 8.1. New PCEP Object

As described in Section 4.1, a new PCEP Object is defined to carry frequency-slot assignment related constraints. IANA is to allocate the following from "PCEP Objects" sub-registry (http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-objects):

| Object Class Value | Name | Object Type | Reference |
|---|---|---|---|
| TDB | SA | 1: Spectrum Assignment | [This.I-D] |

### 8.2. New PCEP TLV: Frequency Slot Selection TLV

As described in Sections 4.2, a new PCEP TLV is defined to indicate spectrum selection constraints. IANA is to allocate this new TLV from the "PCEP TLV Type Indicators" subregistry (http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-tlv-type-indicators).

| Value | Description | Reference |
| --- | --- | --- |
| TBD | Spectrum Selection | [This.I-D] |

### 8.3. New PCEP TLV: Frequency Slot Restriction Constraint TLV

As described in Section 4.3, a new PCEP TLV is defined to indicate wavelength restriction constraints. IANA is to allocate this new TLV from the "PCEP TLV Type Indicators" subregistry (http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-tlv-type-indicators).

| Value | Description | Reference |
| --- | --- | --- |
| TBD | Frequency Slot Restriction Constraint | [This.I-D] |

### 8.4. New PCEP TLV: Spectrum Allocation TLV

As described in Section 5, a new PCEP TLV is defined to indicate the allocation of freq-slots(s) by the PCE in response to a request by the PCC. IANA is to allocate this new TLV from the "PCEP TLV Type Indicators" subregistry (http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-tlv-type-indicators).

| Value | Description | Reference |
| --- | --- | --- |
| TBD | Spectrum Allocation | [This.I-D] |

### 8.5. New No-Path Reasons

As described in Section 4.3, a new bit flag are defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object. This flag, when set, indicates that no feasible route was found that meets all the RSA constraints (e.g., spectrum restriction, signal compatibility, etc.) associated with a RSA path computation request.

IANA is to allocate this new bit flag from the "PCEP NO-PATH-VECTOR TLV Flag Field" subregistry (http://www.iana.org/assignments/pcep/pcep.xhtml#no-path-vector-tlv).

| Bit | Description | Reference |
|---|---|---|
| TBD | No RSA constraints met | [This.I-D] |

### 8.6. New Error-Types and Error-Values

As described in Section 5.1, new PCEP error codes are defined for WSON RWA errors. IANA is to allocate from the ""PCEP-ERROR Object Error Types and Values" sub-registry (http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-error-object).

| Error-Type | Meaning | Error-Value | Reference |
|---|---|---|---|
| TDB | Flexi-Grid RSA Error | 1: Insufficient Memory | [This.I-D] |
| | | 2: RSA computation Not supported | [This.I-D] |

# 9. References

## 9.1. Informative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2863] McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB", RFC 2863, June 2000.

[RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, February 2005.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol", RFC 5440, March 2009.

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

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

[RFC6163] Lee, Y. and Bernstein, G. (Editors), and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, March 2011.

[RFC6566] Y. Lee, G. Bernstein, D. Li, G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSON) with Impairments", RFC 6566, March 2012.

[RFC7420] Koushik, A., E. Stephan, Q. Zhao, D. King, and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, December 2014.

[RFC7446] Y. Lee, G. Bernstein. (Editors), "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", RFC 7446, February 2015.

[RFC7449] Lee, Y., et. al., "PCEP Requirements for WSON Routing and Wavelength Assignment", RFC 7449, February 2015.

## 9.2. Normative References

[PCEP-GMPLS] Margaria, et al., "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions, work in progress.

[RFC5420] Farrel, A. "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC5420, February 2009.

[RFC5521] Oki, E, T. Takeda, and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, April 2009.

[RFC6205] Tomohiro, O. and D. Li, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", RFC 6205, January, 2011.

[RFC7570] Margaria, et al., "Label Switched Path (LSP) Attribute in the Explicit Route Object (ERO)", RFC 7570, July 2015.

[RFC7689] Bernstein et al, "Signaling Extensions for Wavelength Switched Optical Networks", RFC 7689, November 2015.

[RFC7688] Y. Lee, and G. Bernstein, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", RFC 7688, November 2015.

[RFC7698] O. Gonzalez de Dios, R. Casellas, editors, "Framework and Requirements for GMPLS-Based Control of Flexi-Grid Dense Wavelength Division Multiplexing (DWDM) Networks", RFC 7698, November 2015.

[RFC7581] Bernstein and Lee, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", RFC7581, June 2015.

[RFC7579] Bernstein and Lee, "General Network Element Constraint Encoding for GMPLS Controlled Networks", RFC 7579, June 2015.

[PCEP-WSON] Y. Lee (Ed.), and R. Casellas (Ed.), "PCEP Extension for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-rwa-ext, work in progress.

## 10. Contributors

## Authors' Addresses

Young Lee, Editor
Huawei Technologies

Email: leeyoung@huawei.com

Haomian Zheng
Huawei Technologies

Email: zhenghaomian@huawei.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Email: ramon.casellas@cttc.es

Ricard Vilalta
CTTC
Email: ricard.vilalta@cttc.es

Daniele Ceccarelli
Ericsson AB
Gronlandsgatan 21
Kista - Stockholm
Email: daniele.ceccarelli@ericsson.com

Francesco Lazzeri
Ericsson
Via Melen 77
Genova - Italy
Email: francesco.lazzeri@ericsson.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2019

C. Li
M. Chen
D. Dhody
Huawei Technologies
W. Cheng
China Mobile
Z. Li
J. Dong
Huawei Technologies
R. Gandhi
Cisco Systems, Inc.

October 22, 2018

# PCEP Extension for Segment Routing (SR) Bidirectional Associated Paths
draft-li-pce-sr-bidir-path-02

## Abstract

The Path Computation Element Communication Protocol (PCEP) provides mechanisms for Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients (PCCs) requests. The Stateful PCE extensions allow stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Label Switched Paths (LSPs) using PCEP. Furthermore, PCEP can be used for computing paths in SR networks.

This document defines PCEP extensions for grouping two reverse unidirectional SR Paths into an Associated Bidirectional SR path when using a Stateful PCE for both PCE-Initiated and PCC-Initiated LSPs as well as when using a Stateless PCE.

## Status of This Memo

Copyright Notice



Table of Contents



## 1. Introduction

Segment routing (SR) [RFC8402] leverages the source routing and tunneling paradigms. SR supports to steer packets into an explicit forwarding path at the ingress node.

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalized MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

[I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP) [RFC5440] for SR networks, that allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request, report or delegate SR paths. [I-D.negi-pce-segment-routing-ipv6] extend PCEP to support SR for IPv6 data plane.

[I-D.ietf-pce-association-group] introduces a generic mechanism to create a grouping of LSPs which can then be used to define associations between a set of LSPs and/or a set of attributes, for example primary and secondary LSP associations, and is equally applicable to the active and passive modes of a Stateful PCE [RFC8231] or a stateless PCE [RFC5440].

Currently, SR network only supports unidirectional path, but the bidirectional SR path is required in some scenarios, for example, mobile backhaul transport network. The requirement of SR bidirectional path is specified in [I-D.cheng-spring-mpls-path-segment].

[I-D.ietf-pce-association-bidir] defines PCEP extensions for grouping two reverse unidirectional MPLS TE LSPs into an Associated Bidirectional LSP when using a Stateful PCE for both PCE-Initiated and PCC-Initiated LSPs as well as when using a Stateless PCE.

This document extends the bidirectional association to segment routing by specifying PCEP extensions for grouping two reverse unidirectional SR paths into a bidirectional SR path.

[I-D.ietf-pce-association-bidir] specify the Double-sided Bidirectional procedure, where the PCE creates the association and provisions at the both ends, the RSVP-TE does the signaling to the egress the status of the forward LSP and the ingress about the reverse LSP. Thus, the both ends learn both the LSPs forming the bidirectional association. In case of SR, to support the bidirectional use-case, this is done via the PCEP protocol itself as described in Section 3.1. This is done so that both ends are aware of the Path Segment used by each of the unidirectional LSP, as well as the status, the ERO etc.

[I-D.li-pce-sr-path-segment] defines a procedure for Path Segment in PCEP for SR by defining the PATH-SEGMENT TLV. The Path Segment can be a Path Segment in SR-MPLS [I-D.cheng-spring-mpls-path-segment], or a Path Segment in SRv6 [I-D.li-spring-srv6-path-segment], or other IDs that can identify an SR path. The PATH-SEGMENT TLV SHOULD be included for associated bidirectional SR paths.

## 2. Terminology

This document makes use of the terms defined in [I-D.ietf-pce-segment-routing]. The reader is assumed to be familiar with the terminology defined in [RFC5440], [RFC8231], [RFC8281], [I-D.ietf-pce-association-group] and [I-D.ietf-pce-association-bidir].

### 2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 3. PCEP Extension for Bidirectional SR Path

As per [I-D.ietf-pce-association-group], LSPs are associated by adding them to a common association group. [I-D.ietf-pce-association-bidir] specifies PCEP extensions for grouping two reverse unidirectional MPLS-TE LSPs into an Associated Bidirectional LSP for both single-sided and double-sided initiation cases by defining two new Bidirectional LSP Association Groups.

This document extends the procedure for SR bidirectional associated paths by defining a new bidirectional association type (i.e. Double-sided Bidirectional SR Path Association Group). The document further describes the mechanism of associating two unidirectional SR path into a bidirectional SR path. [I-D.li-pce-sr-path-segment] defines a procedure for Path Segment in PCEP for SR by defining the PATH-SEGMENT TLV. The bidirectional SR path can also use the PATH-SEGMENT TLV.

Note that a new association type is created by this document to create new procedures applicable to SR-path (and are quite different than the RSVP-TE bidirectional association groups).

### 3.1. Double-sided Bidirectional SR Path Association Group Object

As defined in [I-D.ietf-pce-association-bidir], two LSPs are associated as a bidirectional MPLS-TE LSP by a common bidirectional LSP association group. For associating two SR paths, this document defines a new association group called 'Double-sided Bidirectional SR Path Association Group' as follows:

- o Association Type (TBD1 to be assigned by IANA) = Double-sided Bidirectional SR Path Association Group

Similar to other bidirectional associations, this Association Type is operator-configured in nature and statically created by the operator on the PCEP peers. The paths belonging to this association is conveyed via PCEP messages to the PCEP peer. Operator-configured Association Range TLV [I-D.ietf-pce-association-group] MUST NOT be sent for these Association Types, and MUST be ignored, so that the entire range of association ID can be used for them. The handling of the Association ID, Association Source, optional Global Association Source and optional Extended Association ID in this association are set in the same way as [I-D.ietf-pce-association-bidir].

A member of the Double-sided Bidirectional SR Path Association Group can take the role of a forward or reverse SR path and follows the rules similar to the rules defined in [I-D.ietf-pce-association-bidir] for LSPs.

- o An SR path (forward or reverse) can not be part of more than one Double-sided Bidirectional SR Path Association Group.

- o The endpoints of the SR paths in this associations cannot be different.

For describing the SR paths in this association group, such as direction and co-routed information, this association group reuses the Bidirectional LSP Association Group TLV defined in [I-D.ietf-pce-association-bidir]. All fields and processing rules are as per [I-D.ietf-pce-association-bidir].

## 4. Bidirectional Flag

As defined in [RFC5440], the B-flag in RP object MUST be set when the PCC specifies that the path computation request relates to a bidirectional TE LSP. In this document, the B-flag also MUST be set when the PCC specifies that the path computation request relates to a bidirectional SR path. When a stateful PCE initiates or updates a bidirectional SR paths including LSPs and SR paths, the B-flag in SRP object [I-D.ietf-pce-pcep-stateful-pce-gmpls] may be set as well.

## 5. Procedures of Bidirectional Path Computation

Two unidirectional SR paths can be associated by the association group object as specified in [I-D.ietf-pce-association-group]. A bidirectional LSP association group object is defined in [I-D.ietf-pce-association-bidir] (for MPLS-TE). This documents extends the mechanism for bidirectional SR paths. Two SR paths can be associated together by including the Bidirectional SR Path Association Group in the PCEP messages. The PATH-SEGMENT TLV [I-D.li-pce-sr-path-segment] SHOULD also be included in the LSP object for these SR paths to support required use-cases.

There is also a need to include the reverse direction path in the PCEP messages, to do this the PCE SHOULD inform the reverse SR path to the ingress PCC and vice versa. To achieve this a PCInitiate message for the reverse SR path is sent to the ingress PCC and a PCInitiate message for the forward SR path is sent to the egress PCC (with the same association group). These PCInitiate message MUST NOT trigger initiation of SR paths. The information of reverse direction path can be used for several scenarios, such as directed BFD [I-D.ietf-mpls-bfd-directed].

### 5.1. PCE Initiated SR Paths

As specified in [I-D.ietf-pce-association-group] Bidirectional SR Association Group can be created by a Stateful PCE.

- o Stateful PCE can create and update the forward and reverse SR path independently for Double-sided Bidirectional SR Path Association Groups.

- o Stateful PCE can establish and remove the association relationship on a per SR path basis.

- o Stateful PCE can create and update the SR path and the association on a PCC via PCInitiate and PCUpd messages, respectively, using the procedures described in [I-D.ietf-pce-association-group].

- o The PATH-SEGMENT TLV SHOULD be included for each SR path in the LSP object.

- o The opposite direction SR path (LSP2(R) at S, LSP1(F) at D ) SHOULD be informed via PCInitiate message with the matching association group.

```
                             +-----+
                             | PCE |
                             +-----+
   PCUpd/PCInitiate         /     \         PCUpd/PCInitiate
   Tunnel 1 (F)            /       \        Tunnel 2 (R)
   (LSP1 (F), LSP2 (R))   /         \       (LSP2 (R), LSP1 (F))
   Assoc#1               /           \      Assoc#1
                        /             \
                       v               v
                  +-----+    LSP1     +-----+
                  |  S  |------------>|  D  |
                  |     |<------------|     |
                  +-----+    LSP2     +-----+
                        <no signaling>
```

Figure 1: PCE-Initiated Double-sided Bidirectional SR Path

### 5.2. PCC Initiated SR Paths

As specified in [I-D.ietf-pce-association-group] Bidirectional SR Association Group can also be created by a PCC.

- o PCC can create and update the forward and reverse SR paths independently for Double-sided Bidirectional SR Path Association Groups.

- o PCC can establish and remove the association relationship on a per SR path basis.

- o PCC MUST report the change in the association group of an SR path to PCE(s) via PCRpt message.

- o PCC can report the forward and reverse SR paths independently to PCE(s) via PCRpt message.

- o PCC can delegate the forward and reverse SR paths independently to a Stateful PCE, where PCE would control the SR paths.

- o Stateful PCE can update the SR paths in the Double-sided Bidirectional SR Path Association Group via PCUpd message, using the procedures described in [I-D.ietf-pce-association-group].

- o The PATH-SEGMENT TLV MUST be handled as defined in [I-D.li-pce-sr-path-segment].

o The opposite direction SR path (LSP2(R) at S, LSP1(F) at D ) SHOULD be informed via PCInitiate message with the matching association group.

```
                            +-----+
                            | PCE |
                            +-----+
Reports/Delegates:          ^    ^       Reports/Delegates
Tunnel 1 (F)               /      \        Tunnel 2 (R)
(LSP1 (F))                /        \       (LSP2 (R))
                         /          \
                        /            \
                       /              \
                   +-----+    LSP1     +-----+
                   |  S  |------------>|  D  |
                   |     |<------------|     |
                   +-----+    LSP2     +-----+
```

Figure 2a: PCC-Initiated Double-sided Bidirectional SR Path

```
                            +-----+
                            | PCE |
                            +-----+
PCUpd/PCInitiate            /    \          PCUpd/PCInitiate
Tunnel 1 (F)               /      \         Tunnel 2 (R)
(LSP1 (F), LSP2 (R))      /        \        (LSP2 (R), LSP1 (F))
Assoc#1                  /          \       Assoc#1
                        /            \
                       v              v
                  +-----+    LSP1     +-----+
                  |  S  |------------>|  D  |
                  |     |<------------|     |
                  +-----+    LSP2     +-----+
```

Figure 2b: PCC-Initiated Double-sided Bidirectional SR Path
along with opposite direction SR path

## 5.3. Error Handling

The error handling as described in section 5.5 of [I-D.ietf-pce-association-bidir] continue to apply.

The Path Setup Type (PST) MUST be set to SR for the LSP belonging to the 'Double-sided Bidirectional SR Path Association Group', in case a PCEP speaker receives a different PST value, it MUST send an PCErr message with Error-Type = 29 (Early allocation by IANA) (Association

Error) and Error-Value = TBD2 (Bidirectional LSP Association - PST Mismatch).

## 6. IANA Considerations

### 6.1. Association Type

This document defines a new Association Type for the Association Object defined [I-D.ietf-pce-association-group]. IANA is requested to make the assignment of a value for the sub-registry "ASSOCIATION Type Field" (to be created in [I-D.ietf-pce-association-group]), as follows:

| Value | Name | Reference |
|---|---|---|
| TBD1 | Double-sided Bidirectional SR Path Association Group | This document |

### 6.2. PCEP Errors

This document defines new Error value for Error Type 29 (Association Error). IANA is requested to allocate new Error value within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry, as follows:

| Error Type | Description | Reference |
|---|---|---|
| 29 | Association Error | |
| | Error value: TBD2 Bidirectional LSP Association - PST Mismatch | This document |

## 7. Security Considerations

The security considerations described in [RFC5440], [RFC8231], [RFC8281], and [I-D.ietf-pce-segment-routing] apply to the extensions defined in this document as well.

A new Association Type for the Association Object, Double-sided Associated Bidirectional SR Path Association Group is introduced in this document. Additional security considerations related to LSP associations due to a malicious PCEP speaker is described in [I-D.ietf-pce-association-group] and apply to this Association Type. Hence, securing the PCEP session using Transport Layer Security (TLS) [RFC8253] is recommended.

## 8. Acknowledgments

## 9. References

### 9.1. 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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[I-D.ietf-pce-association-group]
Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-06 (work in progress), June 2018.

[I-D.ietf-pce-association-bidir]
Barth, C., Gandhi, R., and B. Wen, "PCEP Extensions for Associated Bidirectional Label Switched Paths (LSPs)", draft-ietf-pce-association-bidir-01 (work in progress), May 2018.

[I-D.ietf-pce-pcep-stateful-pce-gmpls]
Zhang, X., Lee, Y., Zhang, F., Casellas, R., Dios, O., and Z. Ali, "Path Computation Element (PCE) Protocol Extensions for Stateful PCE Usage in GMPLS-controlled Networks", draft-ietf-pce-pcep-stateful-pce-gmpls-08 (work in progress), February 2018.

[I-D.negi-pce-segment-routing-ipv6]
Negi, M., Dhody, D., Sivabalan, S., and P. Kaladharan, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-negi-pce-segment-routing-ipv6-03 (work in progress), October 2018.

[I-D.li-pce-sr-path-segment]
Li, C., Chen, M., Dhody, D., Cheng, W., Dong, J., Li, Z., and R. Gandhi, "Path Computation Element Communication Protocol (PCEP) Extension for Path Identification in Segment Routing (SR)", draft-li-pce-sr-path-segment-02 (work in progress), September 2018.

[I-D.li-spring-srv6-path-segment]
Li, C., Chen, M., Dhody, D., Li, Z., Dong, J., and R. Gandhi, "Path Segment for SRv6 (Segment Routing in IPv6)", draft-li-spring-srv6-path-segment-00 (work in progress), October 2018.

## 9.2. Informative References

[RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <https://www.rfc-editor.org/info/rfc4657>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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-14 (work in progress), October 2018.

[I-D.ietf-mpls-bfd-directed]
Mirsky, G., Tantsura, J., Varlashkin, I., and M. Chen, "Bidirectional Forwarding Detection (BFD) Directed Return Path", draft-ietf-mpls-bfd-directed-10 (work in progress), September 2018.

[I-D.cheng-spring-mpls-path-segment]
Cheng, W., Wang, L., Li, H., Chen, M., Gandhi, R., Zigler, R., and S. Zhan, "Path Segment in MPLS Based Segment Routing Network", draft-cheng-spring-mpls-path-segment-03 (work in progress), October 2018.

Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing  100095
China

Email: chengli13@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing  100095
China

Email: Mach.chen@huawei.com

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka  560066
India

Email: dhruv.ietf@gmail.com

Weiqiang Cheng
China Mobile
China

Email: chengweiqiang@chinamobile.com

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing  100095
China

Email: lizhenbin@huawei.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing  100095
China

Email: jie.dong@huawei.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

PCE Working Group C. Li
Internet-Draft M. Chen
Intended status: Standards Track D. Dhody
Expires: April 25, 2019 Huawei Technologies
W. Cheng
China Mobile
J. Dong
Z. Li
Huawei Technologies
R. Gandhi
Cisco Systems, Inc.
October 22, 2018

# Path Computation Element Communication Protocol (PCEP) Extension for Path Segment in Segment Routing (SR)
draft-li-pce-sr-path-segment-03

Abstract

The Path Computation Element (PCE) provides path computation functions in support of traffic engineering in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Path identification is needed for several use cases such as performance measurement in Segment Routing (SR) network. This document specifies extensions to the Path Computation Element Protocol (PCEP) to support requesting, replying, reporting and updating the Path Segment ID (Path SID) between PCEP speakers.

Table of Contents

## 1. Introduction

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalzied MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment routing (SR) [RFC8402] leverages the source routing and tunneling paradigms and supports steering packets into an explicit forwarding path at the ingress node.

An SR path needs to be identified in some use cases such as performance measurement. For identifying an SR path, [I-D.cheng-spring-mpls-path-segment] introduces a new segment that is referred to as Path Segment.

[I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP) [RFC5440] for SR networks, that allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request, report or delegate SR paths. [I-D.negi-pce-segment-routing-ipv6] extend PCEP to support SR paths for IPv6 data plane.

[I-D.zhao-pce-pcep-extension-pce-controller-sr] specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR SID distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network.

This document specifies a mechanism to carry the SR path identification information in PCEP messages [RFC5440] [RFC8231] [RFC8281]. The SR path identifier can be a Path Segment in SR-MPLS [I-D.cheng-spring-mpls-path-segment], or a Path Segment in SRv6 [I-D.li-spring-srv6-path-segment] or other IDs that can identify an SR path. This document also extends the PCECC-SR mechanism to inform the Path Segment to the egress PCC.

## 2. Terminology

This memo makes use of the terms defined in [RFC4655], [I-D.ietf-pce-segment-routing], and [RFC8402].

### 2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 3. Overview of Path Segment Extensions in PCEP

This document specifies a mechanism of encoding (and allocating) Path Segment in PCEP extensions. For supporting Path Segment in PCEP, several TLVs and flags are defined. The formats of the objects and TLVs are described in Section 4. The procedures of Path Segment allocation are described in Section 5.

There are various modes of operations, such as -

- o The Path Segment can be allocated by Egress PCC. The PCE should request the Path Segment from Egress PCC.

- o The PCE can allocate a Path Segment on its own accord and inform the ingress/egress PCC, useful for PCE-initiated LSPs.

- o Ingress PCC can also request PCE to allocate the Path Segment, in this case, the PCE would either allocate and inform the assigned Path Segment to the ingress/egress PCC using PCEP messages, or first request egress PCC for Path Segment and then inform it to the ingress PCC.

The path information to the ingress PCC and PCE is exchanged via an extension to [I-D.ietf-pce-segment-routing] and [I-D.negi-pce-segment-routing-ipv6]. The Path Segment information to the egress PCC can be informed via an extension to the PCECC-SR procedures [I-D.zhao-pce-pcep-extension-pce-controller-sr].

For the PCE to allocate a Path Segment, the PCE SHOULD be aware of the MPLS label space from the PCCs. This is done via mechanism as described in [I-D.li-pce-controlled-id-space]. Otherwise, the PCE should request the egress PCC for Path Segment allocation.

# 4. Objects and TLVs

## 4.1. The OPEN Object

### 4.1.1. The SR PCE Capability sub-TLV

[I-D.ietf-pce-segment-routing] defined a new Path Setup Type (PST) and SR-PCE-CAPABILITY sub-TLV for SR. PCEP speakers use this sub-TLV to exchange information about their SR capability. The TLV defines a Flags field that includes one bit (L-flag) to indicate Local Significance [I-D.ietf-pce-segment-routing].

This document adds an additional flag for Path Segment allocation, as follows -

P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SR path identification (Path Segment, as per [I-D.cheng-spring-mpls-path-segment]).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type=26           |            Length=4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Reserved              |  Flags |P|N|L|      MSD       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: P-flag in SR-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and can be removed at the time of publication.

#### 4.1.2. The SRv6 PCE Capability sub-TLV

[I-D.negi-pce-segment-routing-ipv6] defined a new Path Setup Type (PST) and SRv6-PCE-CAPABILITY sub-TLV for SRv6. PCEP speakers use this sub-TLV to exchange information about their SRv6 capability. The TLV includes a Flags field and one bit (L-flag) was allocated in [I-D.negi-pce-segment-routing-ipv6].

This document adds an additional flag for Path Segment allocation, as follows -

P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SRv6 path identification.(Path Segment, as per [I-D.li-spring-srv6-path-segment]).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Type=TBD1          |            Length=4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     max-SL    |   Reserved    |             Flags         |P|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: P-flag in SRv6-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and can be removed at the time of publication.

#### 4.1.3. PCECC-CAPABILITY sub-TLV

Along with the SR sub-TLVs, the PCECC Capability as per [I-D.zhao-pce-pcep-extension-pce-controller-sr] should be advertised if the PCE allocates the Path Segment and acts as a Central Controller that manages the Label space.

The PCECC Capability should also be advertised on the egress PCEP session, along with the SR sub-TLVs. This is needed to ensure that the PCE can use the PCECC objects/mechanism to request/inform the egress PCC of the Path Segment as described in this document.

### 4.2. LSP Object

The LSP Object is defined in Section 7.3 of [RFC8231]. This document adds the following flags to the LSP Object:

P (Path Segment Allocation bit): If the bit is set to 1, it indicates that the Path Segment needs to be allocated by the PCE for this LSP. A PCC would set this bit to 1 to request for allocation of Path Segment by the PCE in the PCReq or PCRpt message. A PCE would also set this bit to 1 to indicate that the Path Segment is allocated by PCE and encoded in the PCRep, PCUpd or PCInitiate message (the PATH-SEGMENT TLV MUST be present in LSP object). Further, a PCE would set this bit to 0 to indicate that the Path Segment should be allocated by the PCC as described in Section 5.1.1.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                PLSP-ID                | Flag|P|C|  O  |A|R|S|D|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                        TLVs                                  //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: P-flag in LSP Object

The figure is included for the ease of the reader and can be removed at the time of publication.

#### 4.2.1. Path Segment TLV

The PATH-SEGMENT TLV is an optional TLV for use in the LSP Object for Path Segment allocation. The type of this TLV is to be allocated by IANA (TBA4). The format is shown below.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Type             |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       ST      |  Flag       |L|           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               (Variable length) Path Segment                  ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: The PATH-SEGMENT TLV Format

The type (16-bit) of the TLV is TBA4 (to be allocated by IANA). The length (16-bit) has a fixed value of 8 octets. The value contains the following fields:

ST (The Segment type - 8 bits): The ST field specifies the type of the Path Segment field, which carries a Path Segment corresponding to the SR path.

* 0: MPLS Path Segment, which is an MPLS label as defined in [I-D.cheng-spring-mpls-path-segment]. The PST type MUST be set to SR (MPLS).

* 1: SRv6 Path Segment, which is a 128 bit IPv6 address as defined in [I-D.li-spring-srv6-path-segment]. The PST type MUST be set to SRv6.

Flags (8 bits): Two flags are currently defined:

* L-Bit (Local/Global - 1 bit): If set, then the Path Segment carried by the PATH-SEGMENT TLV has local significance. If not set, then the Path Segment carried by this TLV has global significance (i.e. Path Segment is global within an SR domain).

* The unassigned bits MUST be set to 0 and MUST be ignored at receipt.

Reserved (16 bits): MUST be set to 0 and MUST be ignored at receipt.

Path Segment: The Path Segment of an SR path. The Path Segment type is indicated by the ST field. When the ST is 0, it is a MPLS Path Segment [I-D.cheng-spring-mpls-path-segment] in the MPLS label format. When the ST field is 1, it is a 128-bit SRv6 Path Segment as defined in [I-D.li-spring-srv6-path-segment].

In general, only one instance of PATH-SEGMENT TLV will be included in LSP object. If more than one PATH-SEGMENT TLV is included, the first one is processed and others MUST be ignored. Multiple Path Segment allocation for use cases like alternate-making will be considered in future version of this draft.

When the Path Segment allocation is enable, a PATH-SEGMENT TLV should be included in the LSP object.

If the label space is maintained by PCC itself, and the Path Segment is allocated by Egress PCC, then the PCE should request the Path Segment from Egress PCC as described in Section 5.1.1. In this case, the PCE should send a PCUpdate or PCInitiate message to the egress PCC to request the Path Segment. The P-flag in LSP should be unset in this case.

If the PCC requests the Path Segment to be allocated by the PCE, P flag in LSP object is set to 1 and PATH-SEGMENT TLV MAY be skipped. After the PCE has allocated a Path Segment, it MUST include the PATH-SEGMENT TLV in a LSP object.

If the PCE allocated the Path Segment on its own accord, a PATH-SEGMENT TLV MUST be included in a LSP object.

If a PCEP node does not recognize the PATH-SEGMENT TLV, it would behave in accordance with [RFC5440] and ignore the TLV. If a PCEP node recognizes the TLV but does not support the TLV, it MUST send PCErr with Error-Type = 2 (Capability not supported).

## 4.3. FEC Object

The FEC Object [I-D.zhao-pce-pcep-extension-pce-controller-sr] is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message for the PCECC-SR operations. The PCE MUST inform the Path Identification information to the Egress PCC. To do this, this document extends the procedures of [I-D.zhao-pce-pcep-extension-pce-controller-sr] by defining a new FEC object type for Path.

FEC Object-Type is TBA6 'Path'.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                           TLV(s)                            //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: The path FEC object Format

One or more following TLV(s) are allowed in the 'path' FEC object -

- o SYMBOLIC-PATH-NAME TLV: As defined in [RFC8231], it is a human-readable string that identifies an LSP in the network.
- o LSP-IDENTIFIERS TLVs: As defined in [RFC8231], it is optional for SR, but could be used to encode the source, destination and other identification information for the path.

o SPEAKER-ENTITY-ID TLV: As defined in [RFC8232], a unique identifier for the PCEP speaker, it is used to identify the Ingress PCC.

Either SYMBOLIC-PATH-NAME TLV or LSP-IDENTIFIERS TLV MUST be included. SPEAKER-ENTITY-ID TLV is optional. Only one instance of each TLV is processed, if more than one TLV of each type is included, the first one is processed and others MUST be ignored.

## 4.4. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the forwarding instructions is defined in [I-D.zhao-pce-pcep-extension-for-pce-controller]. Further [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object type for SR.

The Path Segment information is encoded directly in the CCI SR object. The Path Segment TLV as described in the Section 4.2.1, MUST also be included in the CCI SR object as the TLV (as it includes additional information regarding the Path Segment identifier).

This document adds the following flags to the CCI Object:

o C (PCC Allocation bit): If the bit is set to 1, it indicates that the allocation needs to be done by the PCC for this central controller instruction. A PCE set this bit to request the PCC to make an allocation from its SR label space. A PCC would set this bit to indicate that it has allocated the CC-ID and report it to the PCE.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            CC-ID                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      MT-ID    |  Algorithm    |        Flags    |C|N|E|V|L|O|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                 SID/Label/Index (variable)                  //
+---------------------------------------------------------------+
|                                                               |
//                       Optional TLV                          //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: The CCI object for SR

(Editor's Note - An update is planned for [I-D.zhao-pce-pcep-extension-pce-controller-sr] in the next revision detailing this procedure, and the above text might move there.)

# 5. Operations

The Path Segment allocation and encoding is as per the stateful PCE operations for segment routing. The procedures are as per the corresponding extensions defined in [I-D.ietf-pce-segment-routing] and [I-D.negi-pce-segment-routing-ipv6] (which are further based on [RFC8231] and [RFC8281]). The additional operations for Path Segment are defined in this section.

To notify (or request) the Path Segment to the Egress PCC, the procedures are as per the PCECC-SR [I-D.zhao-pce-pcep-extension-pce-controller-sr] (which is based on [I-D.zhao-pce-pcep-extension-for-pce-controller]). The additional operations are defined in this section.

## 5.1. PCC Allocated Path Segment

### 5.1.1. Egress PCC Allocated Path Segment

As defined in [I-D.cheng-spring-mpls-path-segment], a Path Segment can be allocated by the egress PCC. In this case, the label space may be maintained on the PCC itself.

On receiving a stateful path computation request with Path Segment allocation request from an ingress PCC, or by initiating or updating an LSP with Path Segment actively, a PCE can request the egress PCC to allocate a Path Segment. This is needed if the PCE does not control the Path Segment allocation for the egress PCC or the label space is maintained by the egress PCC itself.

The mechanism of Path Segment request and reply may be achieved by using PCInitiate and PCUpd message as described in this section.

#### 5.1.1.1. Using CCI and FEC objects (PCECC)

The PCE can request the egress to allocate the Path Segment using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. The C flag in the CCI object is set to 1 and the CC-ID is set to a special value of 0x0000 to indicate that the allocation needs to be done by the PCC. The PATH-SEGMENT TLV is also be included in CCI object along with the FEC object identifying the SR-Path. The egress PCC would allocate the Path Segment and would report to the PCE using the PCRpt message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr] with the allocated Path Segment in the CC-ID field as well as in the PATH-SEGMENT TLV.

(Editor's Note - An update is planned for [I-D.zhao-pce-pcep-extension-pce-controller-sr] in the next revision detailing this procedure)

If the value of CC-ID/Path Segment is 0 and the C flag is set, it indicates that the PCE is requesting a Path Segment for this LSP. If the CC-ID/Path Segment is set to a value 'n' and the C flag is set in the CCI object, it indicates that the PCE requests a specific value 'n' of Path Segment. If the Path Segment is allocated successfully, the egress PCC should report the Path Segment via PCRpt message with the CCI object along with the PATH-SEGMENT TLV. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment in CCI object is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

Once the PCE receives the PCRpt message with the CCI object, it can obtain the Path Segment information from the egress PCC and then update the path with Path Segment or reply to the ingress PCC, the path information with Path Segment.

If the SR-Path is setup the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, as described in [I-D.ietf-pce-segment-routing], an PCErr message will be sent back to the PCE. The PCE MUST request the withdraw of the Path Segment allocation by sending a PCInitiate message to remove the central controller instruction as per [I-D.zhao-pce-pcep-extension-pce-controller-sr]. When the LSP is deleted or the Path Segment is removed, the PCE should synchronize with the egress PCC.

If the egress PCC wishes to withdraw or modify a previously reported Path Segment value, it MUST send a PCRpt message without any PATH-SEGMENT TLV or with the PATH-SEGMENT TLV containing the new Path Segment respectively in the CCI object. The PCE would further trigger the removal of the central controller instruction as per [I-D.zhao-pce-pcep-extension-pce-controller-sr].

If a PCE wishes to modify a previously requested Path Segment value, it MUST send a new PCInitiate message with an allocation request CC-ID/PATH-SEGMENT TLV containing the new Path Segment value and C flag is set. The PCE should trigger the removal of the older Path Segment next as per [I-D.zhao-pce-pcep-extension-pce-controller-sr].

```
               Ingress                                   Egress
               +-+-+                +-+-+                +-+-+
               |PCC|                |PCE|                |PCC|
               +-+-+                +-+-+                +-+-+
   1) LSP State  | ----  PCRpt ---->  |                     |
      Delegate   |     Delegate=1     |                     |
                 |     P=1            |2) PCE update        |
                 |                    |   the LSP-DB and    |
                 |                    |   request Path SID  |
                 |                    |                     |
                 |                    | --- PCInitiate ---> | Egress
                 |                    |      CC-ID=0        | allocates
                 |                    |      FEC=Path       | a Path-SID
                 |                    |                     | from its
                 |                    | <----- PCRpt ------ | space
                 |                    |       CC-ID=        |
                 |                    |       Path SID      |
                 |                    |                     |
                 |<----  PCUpd ----   |3)Paths update with  |
                 |  PATH-SEGMENT TLV  |  Path SID            |
                 |                    |                     |
   4) LSP State  | -----  PCRpt --->  |                     |
      Report     |                    |                     |
                 |                    |                     |
```

Figure 7: Egress PCC Allocated Path Segment

#### 5.1.1.2. Using LSP objects (PCEP-SR)

The PATH-SEGMENT TLV MUST be included in an LSP object in the PCInitiate message sent from the PCE to the egress to request path identification allocation by the egress PCC. The P flag in LSP object MUST be set to 0. This PCInitiate message to egress PCC would be the similar to the one sent to ingress PCC as per [I-D.ietf-pce-segment-routing], but the egress PCC would only allocate the Path Segment and would not trigger the initiation/update operation.

If the value of Path Segment is 0x0 it indicates that the PCE is requesting a Path Segment for this LSP. If the Path Segment is set to a value ’n’ and the P flag is unset in the LSP object, it indicates that the PCE requests a specific value ’n’ of Path Segment. If the Path Segment is allocated successfully, the egress PCC should report the Path Segment via PCRpt message with PATH-SEGMENT TLV in LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7

("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

Once the PCE receives the PCRpt message, it can obtain the Path Segment information from the egress PCC and then update the path with Path Segment or reply to the ingress PCC, the path information with Path Segment.

If the SR-Path is setup the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, as described in [I-D.ietf-pce-segment-routing], an PCErr message will be sent back to the PCE. The PCE MUST request the withdraw of the Path Segment allocation by sending a PCUpd message to remove the LSP and associated Path Segment by setting the R flag in the SRP object. When the LSP is deleted or the Path Segment is removed, the PCE should send a PCUpd message to synchronize with the egress PCC.

If the egress PCC wishes to withdraw or modify a previously reported Path Segment value, it MUST send a PCRpt message without any PATH-SEGMENT TLV or with the PATH-SEGMENT TLV containing the new Path Segment respectively.

If a PCE wishes to modify a previously requested Path Segment value, it MUST send a PCUpd message with PATH-SEGMENT TLV containing the new Path Segment value and P flag is LSP object would be unset. Absence of the PATH-SEGMENT TLV in PCUpd message means that the PCE wishes to withdraw the Path Segment.

If a PCC receives a valid Path Segment value from a PCE which is different than the current Path Segment, it MUST try to allocate the new value. If the new Path Segment is successfully allocated, the PCC MUST report the new value to the PCE. Otherwise, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

```
               Ingress                                    Egress
               +-+-+                +-+-+                 +-+-+
               |PCC|                |PCE|                 |PCC|
               +-+-+                +-+-+                 +-+-+
   1) LSP State  | ----  PCRpt ---->  |                     |
      Delegate   |     Delegate=1     |                     |
                 |     P=1            |2) PCE update        |
                 |                    |   the LSP-DB and    |
                 |                    |   request Path SID  |
                 |                    |                     |
                 |                    | --- PCInitiate ---> | Egress
                 |                    |      PATH-SEGMENT   | allocates
                 |                    |      TLV in LSP     | a Path-SID
                 |                    |                     | from its
                 |                    | <----- PCRpt ------ | space
                 |                    |       Path SID      |
                 |                    |                     |
                 |<----  PCUpd ----   |3)Paths update with  |
                 |  PATH-SEGMENT TLV  |  Path SID           |
                 |                    |                     |
   4) LSP State  | -----  PCRpt --->  |                     |
      Report     |                    |                     |
                 |                    |                     |
```

Figure 8: Egress PCC Allocated Path Segment

## 5.2. PCE Allocated Path Segment

### 5.2.1. PCE Controlled Label Spaces Advertisement

For allocating the Path Segments to SR paths by the PCEs, the PCE controlled label space MUST be known at PCEs via configurations or any other mechanism. The PCE controlled label spaces MAY be advertised as described in [I-D.li-pce-controlled-id-space].

### 5.2.2. Ingress PCC request Path Segment to PCE

The ingress PCC could request the Path Segment to be allocated by the PCE via PCRpt message as per [RFC8231]. The delegate flag (D-flag) MUST also be set for this LSP. Also, the P-flag in the LSP object MUST be set.

If the ingress requests for an expected value of path segment, then a PATH-SEGMENT TLV MUST be included with the expected value. Else, there is no need to include a PATH-SEGMENT TLV in the LSP object.

If the Path Segment is allocated successfully, the PCE would further respond to Ingress PCC with PCUpd message as per [RFC8231] and MUST include the PATH-SEGMENT TLV in a LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

The active PCE would allocate the Path Segment as per the PATH-SEGMENT flags and in case PATH-SEGMENT is not included, the PCE MUST act based on the local policy.

The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

```
                     Ingress                                    Egress
                     +-+-+                +-+-+                 +-+-+
                     |PCC|                |PCE|                 |PCC|
                     +-+-+                +-+-+                 +-+-+
   1) LSP State        | ----  PCRpt ---->  |                     |
      Delegate         |     Delegate=1     |                     |
                       |     P=1            |2) PCE update        |
                       |                    |   the LSP-DB and    |
                       |                    |   allocate Path SID |
                       |<----  PCUpd ----   |3)Paths update with  |
                       |  PATH-SEGMENT TLV  |  Path SID           |
                       |                    |                     |
   4) LSP State Report | -----  PCRpt --->  |                     |
                       |                    |                     |
                       |5) PCE informs the  | --- PCInitiate ---> |
                       |  Path SID to Egress|     FEC=Path        |
                       |                    |                     |
                       |                    | <-------- PCRpt --- |
                       |                    |                     |
```

Figure 9: Ingress PCC request Path Segment to PCE

### 5.2.3. PCE allocated Path Segment on its own

The PCE could allocate the Path Segment on its own for a PCE-Initiated (or delegated LSP). The allocated Path Segment needs to be informed to the Ingress and Egress PCC. The PCE would use the PCInitiate message [RFC8281] or PCUpd message [RFC8231] towards the Ingress PCC and MUST include the PATH-SEGMENT TLV in the LSP object. The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

```
                     Ingress                                Egress
                     +-+-+                +-+-+             +-+-+
                     |PCC|                |PCE|             |PCC|
                     +-+-+                +-+-+             +-+-+
                       |                    |                    |
                       | <--PCInitiate---   |1)Initiate LSP with |
                       | PATH-SEGMENT TLV   |  Path SID          |
                       |                    |                    |
    2)LSP delegation   |---PCRpt, D=1--->   | (Confirm)          |
                       |                    |                    |
                       |3) PCE informs the  | --- PCInitiate --->|
                       |  Path SID to Egress|     FEC=Path       |
                       |                    |                    |
                       |                    | <-------- PCRpt ---|
                       |                    |                    |
```

Figure 10: PCE allocated Path Segment on its own

## 6. Dataplane Considerations

As described in [I-D.cheng-spring-mpls-path-segment], in an SR-MPLS network, when a packet is transmitted along an SR path, the labels in the MPLS label stack will be swapped or popped. So that no label or only the last label may be left in the MPLS label stack when the packet reaches the egress node. Thus, the egress node cannot determine from which SR path the packet comes. For this reason, it introduces the Path Segment.

Apart from allocation and encoding of the Path Segment (described in this document) for the LSP, it would also be included in the SID/ Label stack of the LSP (usually for processing by the egress). To support this, the Path Segment MAY also be a part of SR-ERO as prepared by the PCE as per [I-D.ietf-pce-segment-routing]. The PCC MAY also include the Path Segment while preparing the label stack based on the local policy and use-case.

## 7. IANA Considerations

### 7.1. SR PCE Capability Flags

SR PCE Capability TLV is defined in [I-D.ietf-pce-segment-routing], and the registry to manage the Flag field of the SR PCE Capability TLV is requested in [I-D.ietf-pce-segment-routing]. IANA is requested to make the following allocation in the aforementioned registry.

| Bit | Description | Reference |
|---|---|---|
| TBA1 | Path Segment Allocation is supported(P) | This document |

### 7.2. SRv6 PCE Capability Flags

SRv6 PCE Capability TLV is defined in defined in [I-D.negi-pce-segment-routing-ipv6], and the registry to manage the Flag field of the SRv6 PCE Capability Flags is requested in [I-D.negi-pce-segment-routing-ipv6]. IANA is requested to make the following allocation in the aforementioned registry.

| Bit | Description | Reference |
|---|---|---|
| TBA2 | Path Segment Allocation is supported(P) | This document |

### 7.3. New LSP Flag Registry

[RFC8231] defines the LSP object; per that RFC, IANA created a registry to manage the value of the LSP object's Flag field. IANA has allocated a new bit in the "LSP Object Flag Field" subregistry, as follows:

| Bit | Description | Reference |
|---|---|---|
| TBA3 | Request for Path Segment Allocation(P) | This document |

### 7.4. New PCEP TLV

IANA is requested to add the assignment of a new allocation in the existing "PCEP TLV Type Indicators" subregistry as follows:

| Value | Description | Reference |
|---|---|---|
| TBA4 | PATH-SEGMENT TLV | This document |

#### 7.4.1. Path Segment TLV

This document requests that a new subregistry named "PATH-SEGMENT TLV Segment Type (ST) Field" to be created to manage the value of the ST field in the PATH-SEGMENT TLV.

| Value | Description | Reference |
|---|---|---|
| 0 | MPLS Path Segment(MPLS label) | This document |
| 1 | SRv6 Path Segment(IPv6 address) | This document |

Further, this document also requests that a new subregistry named "PATH-SEGMENT TLV Flag Field" to be created to manage the Flag field in the PATH-SEGMENT TLV. New values are assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

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

| Bit | Description | Reference |
|---|---|---|
| 7 | Local Signification(L) | This document |

### 7.5. New CCI Flag Registry

CCI object is defined in defined in [I-D.zhao-pce-pcep-extension-for-pce-controller], further [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object type for SR. and the subregistry to manage the Flag field of the CCI object for SR is requested in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. IANA is requested to make the following allocation in the aforementioned subregistry.

| Bit | Description | Reference |
|---|---|---|
| TBA5 | PCC is requested to allocate resource(C) | This document |

### 7.6. New FEC Type Registry

A new PCEP object called FEC is defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. IANA is requested to allocate a new Object-Type for FEC object in the "PCEP Objects" subregistry.

| Value | Description | Reference |
|---|---|---|
| TBA6 | SR path | This document |

### 7.7. PCEP Error Type and Value

IANA is requested to allocate code-points in the "PCEP-ERROR Object Error Types and Values" subregistry for the following new error-types and error-values:

| Error-Type | Meaning | Reference |
|---|---|---|
| TBA7 | Path SID failure:<br>Error-value = 1<br>Invalid SID<br><br>Error-value = 2<br>Unable to allocate<br>Path SID | This document |

## 8. Security Considerations

TBA

## 9. Acknowledgments

## 10. Contributors

The following people have substantially contributed to this document:

Zafar Ali
Cisco Systems, Inc.
Email: zali@cisco.com

## 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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <https://www.rfc-editor.org/info/rfc8232>.

[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>.

[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-14 (work in progress), October 2018.

[I-D.negi-pce-segment-routing-ipv6]
Negi, M., Dhody, D., Sivabalan, S., and P. Kaladharan, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-negi-pce-segment-routing-ipv6-03 (work in progress), October 2018.

[I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-03 (work in progress), June 2018.

[I-D.zhao-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-zhao-pce-pcep-extension-for-pce-controller-08 (work in progress), June 2018.

[I-D.li-spring-srv6-path-segment]
Li, C., Chen, M., Dhody, D., Li, Z., Dong, J., and R. Gandhi, "Path Segment for SRv6 (Segment Routing in IPv6)", draft-li-spring-srv6-path-segment-00 (work in progress), October 2018.

## 11.2. Informative References

[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>.

[RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <https://www.rfc-editor.org/info/rfc4657>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[I-D.li-pce-controlled-id-space]
Li, C., Chen, M., Dong, J., Li, Z., and D. Dhody, "PCE Controlled ID Space", draft-li-pce-controlled-id-space-00 (work in progress), June 2018.

[I-D.cheng-spring-mpls-path-segment]
Cheng, W., Wang, L., Li, H., Chen, M., Gandhi, R., Zigler, R., and S. Zhan, "Path Segment in MPLS Based Segment Routing Network", draft-cheng-spring-mpls-path-segment-03 (work in progress), October 2018.

## Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: chengli13@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: Mach.chen@huawei.com

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

Email: dhruv.ietf@gmail.com

Weiqiang Cheng
China Mobile
China

Email: chengweiqiang@chinamobile.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: jie.dong@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 20, 2020

C. Li
M. Chen
Huawei Technologies
W. Cheng
China Mobile
J. Dong
Z. Li
Huawei Technologies
R. Gandhi
Cisco Systems, Inc.
Q. Xiong
ZTE Corporation
August 19, 2019

# Path Computation Element Communication Protocol (PCEP) Extension for Path Segment in Segment Routing (SR)

draft-li-pce-sr-path-segment-08

## Abstract

The Path Computation Element (PCE) provides path computation functions in support of traffic engineering in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Path identification is needed for several use cases such as performance measurement in Segment Routing (SR) network. This document specifies extensions to the Path Computation Element Communication Protocol (PCEP) to support requesting, replying, reporting and updating the Path Segment ID (Path SID) between PCEP speakers.

## Status of This Memo

Table of Contents

# 1. Introduction

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalized MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment routing (SR) [RFC8402] leverages the source routing and tunneling paradigms and supports steering packets into an explicit forwarding path at the ingress node.

An SR path needs to be identified in some use cases such as performance measurement. For identifying an SR path, [I-D.ietf-spring-mpls-path-segment] introduces a new segment that is referred to as Path Segment.

[I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP) [RFC5440] for SR networks, that allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request, report or delegate SR paths.

[I-D.zhao-pce-pcep-extension-pce-controller-sr] specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR SID distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network.

This document specifies a mechanism to carry the SR path identification information in PCEP messages [RFC5440] [RFC8231] [RFC8281]. The SR path identifier can be a Path Segment in SR-MPLS [I-D.ietf-spring-mpls-path-segment], or other IDs that can identify an SR path. This document also extends the PCECC-SR mechanism to inform the Path Segment to the egress PCC.

## 2. Terminology

This memo makes use of the terms defined in [RFC4655], [I-D.ietf-pce-segment-routing], and [RFC8402].

### 2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 3. Overview of Path Segment Extensions in PCEP

This document specifies a mechanism of allocating Path Segment and extends PCEP to encode it in PCEP messages. For supporting Path Segment in PCEP, several TLVs and flags are defined. The formats of the objects and TLVs are described in Section 4. The procedures of Path Segment allocation are described in Section 5.

There are various modes of operations, such as -

- o The Path Segment can be allocated by Egress PCC. The PCE should request the Path Segment from Egress PCC.

- o The PCE can allocate a Path Segment on its own accord and inform the ingress/egress PCC, useful for PCE-initiated LSPs.

- o Ingress PCC can also request PCE to allocate the Path Segment, in this case, the PCE would either allocate and inform the assigned Path Segment to the ingress/egress PCC using PCEP messages, or first request egress PCC for Path Segment and then inform it to the ingress PCC.

The path information to the ingress PCC and PCE is exchanged via an extension to [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-segment-routing-ipv6]. The Path Segment information (for SR-MPLS) to the egress PCC can be informed via an extension to the PCECC-SR procedures [I-D.zhao-pce-pcep-extension-pce-controller-sr].

For the PCE to allocate a Path Segment on its own, the PCE needs to be aware of the MPLS label space from the PCCs. This is done via mechanism as described in [I-D.li-pce-controlled-id-space]. Otherwise, the PCE should request the egress PCC for Path Segment allocation.

# 4. Objects and TLVs

## 4.1. The OPEN Object

### 4.1.1. The SR PCE Capability sub-TLV

[I-D.ietf-pce-segment-routing] defined a new Path Setup Type (PST) and SR-PCE-CAPABILITY sub-TLV for SR-MPLS. PCEP speakers use this sub-TLV to exchange information about their SR capability. The TLV defines a Flags field [I-D.ietf-pce-segment-routing].

This document adds an additional flag for Path Segment allocation, as follows -

- o P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SR path identification (Path Segment, as per [I-D.ietf-spring-mpls-path-segment]).

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Type=TBD11            |          Length=4             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Reserved              |   Flags |P|N|X|      MSD      |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: P-flag in SR-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and will be removed at the time of publication.

### 4.1.2. PCECC-CAPABILITY sub-TLV

Along with the SR sub-TLVs, the PCECC Capability as per [I-D.zhao-pce-pcep-extension-pce-controller-sr] should be advertised if the PCE allocates the Path Segment and acts as a Central Controller that manages the Label space.

The PCECC Capability should be advertised on the egress PCEP session, along with the SR sub-TLVs. This is needed to ensure that the PCE can use the PCECC objects/mechanism to request/inform the egress PCC of the Path Segment as described in Section 5.2.

## 4.2. LSP Object

The LSP Object is defined in Section 7.3 of [RFC8231]. This document adds a flag in the LSP Object:

- P (PCE Allocation bit): If the bit is set to 1, it indicates that the PCC requests PCE to make allocations for this LSP. The TLV in LSP object identifies what should be allocated, such as Path Segment or Binding Segment. A PCC would set this bit to 1 and include a PATH-SEGMENT TLV in the LSP object to request for allocation of Path Segment by the PCE in the PCEP message. A PCE would also set this bit to 1 and include a PATH-SEGMENT TLV to indicate that the Path Segment is allocated by PCE and encoded in the PCEP message towards PCC. Further, a PCE would set this bit to 0 and include a PATH-SEGMENT TLV in the LSP object to indicate that the Path Segment should be allocated by the PCC as described in Section 5.1.1.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                PLSP-ID                | Flag|P|C|  O  |A|R|S|D|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //                        TLVs                                 //
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: P-flag in LSP Object

The figure is included for the ease of the reader and will be removed at the time of publication.

### 4.2.1. Path Segment TLV

The PATH-SEGMENT TLV is an optional TLV for use in the LSP Object for Path Segment allocation. The type of this TLV is to be allocated by IANA (TBA4). The format is as shown below.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             Type              |             Length            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |       ST      |  Flag       |L|           Reserved            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    ~              (Variable length) Path Segment                   ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: The PATH-SEGMENT TLV Format

The type (16-bit) of the TLV is TBA4 (to be allocated by IANA). The length (16-bit) has a variable length. The value contains the following fields:

- o ST (The Segment type - 8 bits): The ST field specifies the type of the Path Segment field, which carries a Path Segment corresponding to the SR path.

  * 0: MPLS Path Segment, which is an MPLS label as defined in [I-D.ietf-spring-mpls-path-segment]. The PST type MUST be set to SR (MPLS).

  * 1-255: Reserved for future use.

- o Flags (8 bits): One flag is currently defined:

* L-Bit (Local/Global - 1 bit): If set, then the Path Segment carried by the PATH-SEGMENT TLV has local significance. If not set, then the Path Segment carried by this TLV has global significance (i.e. Path Segment is global within an SR domain).

* The unassigned bits MUST be set to 0 and MUST be ignored at receipt.

o Reserved (16 bits): MUST be set to 0 and MUST be ignored at receipt.

o Path Segment: The Path Segment of an SR path. The Path Segment type is indicated by the ST field. When the ST is 0, it is a MPLS Path Segment [I-D.ietf-spring-mpls-path-segment] in the MPLS label format.

In general, only one instance of PATH-SEGMENT TLV will be included in LSP object. If more than one PATH-SEGMENT TLV is included, the first one is processed and others MUST be ignored. Multiple Path Segment allocation for use cases like alternate-making will be considered in future version of this draft.

When the Path Segment allocation is enabled, a PATH-SEGMENT TLV MUST be included in the LSP object.

If the label space is maintained by PCC itself, and the Path Segment is allocated by Egress PCC, then the PCE should request the Path Segment from Egress PCC as described in Section 5.1.1. In this case, the PCE should send a PCUpdate or PCInitiate message to the egress PCC to request the Path Segment. The P-flag in LSP should be unset in this case.

If a PCEP node does not recognize the PATH-SEGMENT TLV, it would behave in accordance with [RFC5440] and ignore the TLV. If a PCEP node recognizes the TLV but does not support the TLV, it MUST send PCErr with Error-Type = 2 (Capability not supported).

## 4.3. FEC Object

The FEC Object [I-D.zhao-pce-pcep-extension-pce-controller-sr] is used to specify the FEC information and carried within PCInitiate or PCRpt message for the PCECC-SR operations. The PCE MUST inform the Path Identification information to the Egress PCC. To do this, this document extends the procedures of [I-D.zhao-pce-pcep-extension-pce-controller-sr] by defining a new FEC object type for Path.

FEC Object-Type is TBA6 'Path'.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                           TLV(s)                            //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: The path FEC object Format

One or more following TLV(s) are allowed in the 'path' FEC object -

- o SYMBOLIC-PATH-NAME TLV: As defined in [RFC8231], it is a human-readable string that identifies an LSP in the network.
- o LSP-IDENTIFIERS TLVs: As defined in [RFC8231], it is optional for SR, but could be used to encode the source, destination and other identification information for the path.
- o SPEAKER-ENTITY-ID TLV: As defined in [RFC8232], a unique identifier for the PCEP speaker, it is used to identify the Ingress PCC.

Either SYMBOLIC-PATH-NAME TLV or LSP-IDENTIFIERS TLV MUST be included. SPEAKER-ENTITY-ID TLV is optional. Only one instance of each TLV is processed, if more than one TLV of each type is included, the first one is processed and others MUST be ignored.

### 4.4. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the forwarding instructions is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]. Further [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object type for SR.

The Path Segment information is encoded directly in the CCI SR object. The Path Segment TLV as described in the Section 4.2.1, MUST also be included in the CCI SR object as the TLV (as it includes additional information regarding the Path Segment identifier). The C flag in CCI object is used to indicate if the allocation needs to be done by the PCC.

## 5. Operations

The Path Segment allocation and encoding is as per the Stateful PCE operations for segment routing. The procedures are as per the corresponding extensions defined in [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-segment-routing-ipv6] (which are further based on [RFC8231] and [RFC8281]). The additional operations for Path Segment are defined in this section.

To notify (or request) the Path Segment to the Egress PCC, the procedures are as per the PCECC-SR [I-D.zhao-pce-pcep-extension-pce-controller-sr] (which is based on [I-D.ietf-pce-pcep-extension-for-pce-controller]). The additional operations are defined in this section.

### 5.1. Stateful PCE Operation

As defined in [I-D.ietf-spring-mpls-path-segment], a Path Segment can be allocated by the egress PCC. In this case, the label space is maintained on the PCC itself.

This section describes the mechanism of Path Segment allocation by using PCInitiate and PCUpd message in Stateful PCE model.

#### 5.1.1. Ingress PCC-Initiated Path Segment Allocation

The ingress PCC could request the Path Segment to be allocated by the PCE via PCRpt message. The delegate flag (D-flag) MUST also be set for this LSP. Also, the P-flag in the LSP object MUST be set.

On receiving a delegation request with Path Segment allocation request from an ingress PCC, a stateful PCE requests the egress PCC to allocate a Path Segment.

The PATH-SEGMENT TLV MUST be included in an LSP object in the PCInitiate message sent from the PCE to the egress to request Path Segment allocation by the egress PCC. The P flag in LSP object MUST be set to 0. This PCInitiate message to egress PCC would be the similar to the one sent to ingress PCC as per [I-D.ietf-pce-segment-routing], but the egress PCC would only allocate the Path Segment and would not trigger the LSP initiation operation (as it would be the egress for this LSP).

If the value of Path Segment is 0x0, it indicates that the PCE is requesting a Path Segment for this LSP. If the Path Segment is set to a value 'n' and the P flag is unset in the LSP object, it indicates that the PCE requests a specific value 'n' of Path Segment. If the Path Segment is allocated successfully, the egress PCC reports the Path Segment via PCRpt message with PATH-SEGMENT TLV in LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

Once the PCE receives the PCRpt message, it can obtain the Path Segment information from the egress PCC and then update the path with Path Segment by sending PCUpd message to the ingress PCC.

If the Path Segment is updated successfully, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, an PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID") will be sent back to the PCE. The PCE MUST roll back the Path Segment value to the previous value (if any) by sending a PCUpd message to synchronize with the egress PCC.

```
            Ingress                                   Egress
            +-+-+                 +-+-+                +-+-+
            |PCC|                 |PCE|                |PCC|
            +-+-+                 +-+-+                +-+-+
1) LSP State  | ----  PCRpt ----> |                      |
   Delegate   |     Delegate=1    |                      |
              |     P=1           |2) PCE update         |
              |                   |   the LSP-DB and     |
              |                   |   request Path SID   |
              |                   |                      |
              |                   | --- PCInitiate ---> | Egress
              |                   |      PATH-SEGMENT    | allocates
              |                   |      TLV in LSP      | a Path-SID
              |                   |                      | from its
              |                   | <----- PCRpt ------  | space
              |                   |        Path SID      |
              |                   |                      |
              |<----  PCUpd ----  |3)Paths update with   |
              |  PATH-SEGMENT TLV |  Path SID            |
              |                   |                      |
4) LSP State  | -----  PCRpt ---> |                      |
   Report     |                   |                      |
              |                   |                      |
```

Figure 5: Ingress PCC-Initiated Path Segment Allocation

If the ingress PCC wishes to withdraw or modify a previously reported Path Segment value, it MUST send a PCRpt message without any PATH-

SEGMENT TLV or with the PATH-SEGMENT TLV containing the new Path Segment respectively. In this case, the PCE should synchronize with egress PCC via PCUpd message.

The Path Segment MUST be withdrawn when the corresponding LSP is removed. When the LSP is deleted, the PCE MUST request the egress PCC to withdraw the LSP and associated Path Segment via PCInitiate message with the R flag is set in the SRP object.

If an egress PCC receives a valid Path Segment value from a PCE which is different than the current Path Segment, it MUST try to allocate the new value. If the new Path Segment is successfully allocated, the egress PCC MUST report the new value to the PCE. Otherwise, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

#### 5.1.2. PCE Initiated Path Segment Allocation

A stateful PCE also can initiate or update an LSP with Path Segment actively via requesting the egress PCC to allocate a Path Segment.

If a PCE wishes to modify a previously requested Path Segment value or allocate a Path Segment for an PCE-Initiated LSP, it MUST request the egress PCC to allocate a new value by sending a PCUpd message to the egress PCC with PATH-SEGMENT TLV containing the new Path Segment value. Also, the P flag in LSP object is unset. Absence of the PATH-SEGMENT TLV in PCUpd message means that the PCE wishes to withdraw the Path Segment.

The mechanism of requesting Path Segment is as per Section 5.1.1.

Once the PCE receives the PCRpt message, it can obtain the Path Segment information from the egress PCC and then update or initiate an LSP with Path Segment.

If the SR-Path is setup, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, as described in [I-D.ietf-pce-segment-routing], an PCErr message will be sent back to the PCE. The PCE MUST request the egress PCC to withdraw the LSP and associated Path Segment via PCInitiate message with the R flag is set in the SRP object.

If the Path Segment is updated successfully, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, an PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID") will be sent back to the PCE. The PCE MUST roll back the Path Segment value to the previous value (if any) by sending a PCUpd message to synchronize with the egress PCC.

```
               Ingress                                    Egress
               +-+-+                +-+-+                 +-+-+
               |PCC|                |PCE|                 |PCC|
               +-+-+                +-+-+                 +-+-+
   1) LSP State  | ----  PCRpt ---->  |                     |
      Delegate if|     Delegate=1     |                     |
   the LSP exists|                    |2)PCE actively update|
                 |                    |  the LSP-DB and     |
                 |                    |  request Path SID   |
                 |                    |                     |
                 |                    | --- PCInitiate ---> | Egress
                 |                    |      PATH-SEGMENT   | allocates
                 |                    |      TLV in LSP     | a Path-SID
                 |                    |                     | from its
                 |                    | <----- PCRpt ------ | space
                 |                    |       Path SID      |
                 |                    |                     |
                 |<-- PCUpd/PCInit -- |3)Paths update with  |
                 |  PATH-SEGMENT TLV  |  Path SID           |
                 |                    |                     |
   4) LSP State  | -----  PCRpt --->  |                     |
      Report     |                    |                     |
                 |                    |                     |
```

Figure 6: Stateful PCE-Initiated Path Segment Allocation

## 5.2. PCECC Based Operation

### 5.2.1. PCE Controlled Label Spaces Advertisement

For allocating the Path Segments to SR paths by the PCEs, the PCE controlled label space MUST be known at PCEs via configurations or any other mechanisms. The PCE controlled label spaces MAY be advertised as described in [I-D.li-pce-controlled-id-space].

### 5.2.2. PCECC based Path Segment Allocation

#### 5.2.2.1. PCECC-Initiated

The PCE could allocate the Path Segment on its own for a PCE-Initiated (or delegated LSP). The allocated Path Segment needs to be informed to the Ingress and Egress PCC. The PCE would use the PCInitiate message [RFC8281] or PCUpd message [RFC8231] towards the Ingress PCC and MUST include the PATH-SEGMENT TLV in the LSP object.

The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

```
                  Ingress                                    Egress
                  +-+-+                +-+-+                 +-+-+
                  |PCC|                |PCE|                 |PCC|
                  +-+-+                +-+-+                 +-+-+
                    |                    |                     |
                    | <--PCInitiate---   |1)Initiate LSP with  |
                    | PATH-SEGMENT TLV   |  Path SID           |
                    |                    |                     |
 2)LSP delegation   |---PCRpt, D=1--->   | (Confirm)           |
                    |                    |                     |
                    |3) PCE informs the  | --- PCInitiate ---> |
                    |  Path SID to Egress|     FEC=Path        |
                    |                    |                     |
                    |                    | <-------- PCRpt --- |
                    |                    |                     |
```

Figure 7: PCE allocated Path Segment on its own

5.2.2.2. Ingress PCC-Initiated PCECC

The ingress PCC could request the Path Segment to be allocated by the PCE via PCRpt message as per [RFC8231]. The delegate flag (D-flag) MUST also be set for this LSP. Also, the P-flag in the LSP object MUST be set.

A PATH-SEGMENT TLV MUST be included in the LSP object. If the value of Path Segment is 0x0, it indicates that the Ingress PCC is requesting a Path Segment for this LSP. If the Path Segment is set to a value 'n', it indicates that the ingress PCC requests a specific value 'n' of Path Segment.

If the Path Segment is allocated successfully, the PCE would further respond to Ingress PCC with PCUpd message as per [RFC8231] and MUST include the PATH-SEGMENT TLV in a LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

The active PCE would allocate the Path Segment as per the PATH-SEGMENT flags and in case PATH-SEGMENT is not included, the PCE MUST act based on the local policy.

The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

```
                     Ingress                                   Egress
                     +-+-+                +-+-+                +-+-+
                     |PCC|                |PCE|                |PCC|
                     +-+-+                +-+-+                +-+-+
1) LSP State           | ----  PCRpt ----> |                      |
   Delegate            |     Delegate=1    |                      |
                       |     P=1           |2) PCE update         |
                       |                   |   the LSP-DB and     |
                       |                   |   allocate Path SID  |
                       |<----  PCUpd ----  |3)Paths update with   |
                       |  PATH-SEGMENT TLV |  Path SID            |
                       |                   |                      |
4) LSP State Report    | -----  PCRpt ---> |                      |
                       |                   |                      |
                       |5) PCE informs the | --- PCInitiate ---> |
                       |  Path SID to Egress|     FEC=Path         |
                       |                   |                      |
                       |                   | <-------- PCRpt ---  |
                       |                   |                      |
```

Figure 8: Ingress PCC request Path Segment to PCE

## 6. Dataplane Considerations

As described in [I-D.ietf-spring-mpls-path-segment], in an SR-MPLS network, when a packet is transmitted along an SR path, the labels in the MPLS label stack will be swapped or popped. So that no label or only the last label may be left in the MPLS label stack when the packet reaches the egress node. Thus, the egress node cannot determine from which SR path the packet comes. For this reason, it introduces the Path Segment.

Apart from allocation and encoding of the Path Segment (described in this document) for the LSP, it would also be included in the SID/Label stack of the LSP (usually for processing by the egress). To support this, the Path Segment MAY also be a part of SR-ERO as prepared by the PCE as per [I-D.ietf-pce-segment-routing]. The PCC MAY also include the Path Segment while preparing the label stack based on the local policy and use-case.

It is important that the PCE learns the Maximum SID Depth (MSD) that can be imposed at each node/link of a given SR path to ensure that the SID stack depth does not exceed the number of SIDs the node is capable of imposing. As a new type of segment, Path Segment will be inserted in the SID list just like other SIDs. Thus, the PCE needs to consider the affect of Path Segment when computing a LSP with Path Segment allocation.

## 7. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to [RFC7942].

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

### 7.1. Huawei’s Commercial Delivery

The feature is developing based on Huawei VRP8.

o Organization: Huawei

o Implementation: Huawei’s Commercial Delivery implementation based on VRP8.

o Description: The implementation is under development and follows the mechanism as defined in section-5.1.1.

o Maturity Level: Product

o Contact: tanren@huawei.com

### 7.2. ZTE's Commercial Delivery

- o Organization: ZTE
- o Implementation: ZTE's Commercial Delivery implementation based on Rosng v8.
- o Description: The implementation is under development and follows the mechanism as defined in section-5.1.1.
- o Maturity Level: Product
- o Contact: zhan.shuangping@zte.com.cn

## 8. IANA Considerations

### 8.1. SR PCE Capability Flags

SR PCE Capability TLV is defined in [I-D.ietf-pce-segment-routing], and the registry to manage the Flag field of the SR PCE Capability TLV is requested in [I-D.ietf-pce-segment-routing]. IANA is requested to make the following allocation in the "SR Capability Flag Field" sub-registry.

| Bit | Description | Reference |
|---|---|---|
| TBA1 | Path Segment Allocation is supported(P) | This document |

### 8.2. New LSP Flag Registry

[RFC8231] defines the LSP object; per that RFC, IANA created a registry to manage the value of the LSP object's Flag field. IANA has allocated a new bit in the "LSP Object Flag Field" sub-registry, as follows:

| Bit | Description | Reference |
|---|---|---|
| TBA3 | Request for Path Segment Allocation(P) | This document |

### 8.3. New PCEP TLV

IANA is requested to add the assignment of a new allocation in the existing "PCEP TLV Type Indicators" sub-registry as follows:

| Value | Description | Reference |
|---|---|---|
| TBA4 | PATH-SEGMENT TLV | This document |

### 8.3.1. Path Segment TLV

This document requests that a new sub-registry named "PATH-SEGMENT TLV Segment Type (ST) Field" to be created to manage the value of the ST field in the PATH-SEGMENT TLV.

| Value | Description | Reference |
|---|---|---|
| 0 | MPLS Path Segment(MPLS label) | This document |
| 1-255 | Reserved for future use | This document |

Further, this document also requests that a new sub-registry named "PATH-SEGMENT TLV Flag Field" to be created to manage the Flag field in the PATH-SEGMENT TLV. New values are assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

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

| Bit | Description | Reference |
|---|---|---|
| 7 | Local Signification(L) | This document |

## 8.4. New FEC Type Registry

A new PCEP object called FEC is defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. IANA is requested to allocate a new Object-Type for FEC object in the "PCEP Objects" sub-registry.

| Value | Description | Reference |
|---|---|---|
| TBA6 | Path | This document |

### 8.5. PCEP Error Type and Value

IANA is requested to allocate code-points in the "PCEP-ERROR Object Error Types and Values" sub-registry for the following new error-types and error-values:

| Error-Type | Meaning | Reference |
|---|---|---|
| TBA7 | Path SID failure:<br>Error-value = 1<br>Invalid SID<br><br>Error-value = 2<br>Unable to allocate<br>Path SID | This document |

## 9. Security Considerations

The security considerations described in [RFC5440]], [RFC8231], [RFC8281] and [I-D.ietf-pce-segment-routing] are applicable to this specification. No additional security measure is required.

As described [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-pcep-extension-for-pce-controller], SR allows a network controller to instantiate and control paths in the network. A rogue PCE can manipulate Path SID allocations to have impact based on the usage of Path SID such as accounting, bi-directional etc.

Thus, as per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253], as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

## 10. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] apply to PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

### 10.1. Control of Function and Policy

A PCEP implementation SHOULD allow the operator to configure the policy based on which it allocates the Path SID. This includes the Path SID scope.

### 10.2. Information and Data Models

The PCEP YANG module is defined in [I-D.ietf-pce-pcep-yang]. In future, this YANG module should be extended or augmented to provide the following additional information relating to Path SID.

An implementation SHOULD allow the operator to view the Path SID allocated to the LSP as well as Path SID as part of the computed SID list for the SR path.

### 10.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

### 10.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] .

### 10.5. Requirements On Other Protocols

Mechanisms defined in this document do not imply any new requirements on other protocols.

### 10.6. Impact On Network Operations

Mechanisms defined in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] also apply to PCEP extensions defined in this document. Further, the mechanism described in this document can help the operator to request control of the LSPs at a particular PCE.

## 11. Acknowledgments

TBA

## 12. References

### 12.1. 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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <https://www.rfc-editor.org/info/rfc8232>.

[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>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <https://www.rfc-editor.org/info/rfc7525>.

[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>.

[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-16 (work in progress), March 2019.

[I-D.ietf-pce-segment-routing-ipv6]
Negi, M., Li, C., Sivabalan, S., Kaladharan, P., and Y. Zhu, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-ietf-pce-segment-routing-ipv6-02 (work in progress), April 2019.

[I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Negi, M., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-05 (work in progress), July 2019.

[I-D.ietf-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Negi, M., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-ietf-pce-pcep-extension-for-pce-controller-02 (work in progress), July 2019.

[I-D.li-spring-srv6-path-segment]
Li, C., Cheng, W., Chen, M., Dhody, D., Li, Z., Dong, J., and R. Gandhi, "Path Segment for SRv6 (Segment Routing in IPv6)", draft-li-spring-srv6-path-segment-03 (work in progress), August 2019.

## 12.2. Informative References

[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>.

[RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <https://www.rfc-editor.org/info/rfc4657>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <https://www.rfc-editor.org/info/rfc7942>.

[I-D.li-pce-controlled-id-space] Li, C., Chen, M., Dong, J., Li, Z., Wang, A., Cheng, W., and C. Zhou, "PCE Controlled ID Space", draft-li-pce-controlled-id-space-03 (work in progress), June 2019.

[I-D.ietf-spring-mpls-path-segment] Cheng, W., Li, H., Chen, M., Gandhi, R., and R. Zigler, "Path Segment in MPLS Based Segment Routing Network", draft-ietf-spring-mpls-path-segment-00 (work in progress), March 2019.

[I-D.ietf-pce-pcep-yang] Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-12 (work in progress), July 2019.

## Appendix A. Contributors

The following people have substantially contributed to this document:

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

Email: dhruv.ietf@gmail.com

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

## Appendix B. SRv6 extensions

This section would be rolled into the document once the SPRING WG adopts SRv6 path segment.

## B.1.  The SRv6 PCE Capability sub-TLV

[I-D.ietf-pce-segment-routing-ipv6] defined a new Path Setup Type (PST) and SRv6-PCE-CAPABILITY sub-TLV for SRv6.  PCEP speakers use this sub-TLV to exchange information about their SRv6 capability. The TLV includes a Flags field and one bit (L-flag) was allocated in [I-D.ietf-pce-segment-routing-ipv6].

This document adds an additional flag for Path Segment allocation, as follows -

o  P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SRv6 path identification.(Path Segment, as per [I-D.li-spring-srv6-path-segment]).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Type=TBD1          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Reserved           |             Flags         |P|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   MSD-Type    |  MSD-Value    |   MSD-Type    |  MSD-Value    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                              ...                            //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   MSD-Type    |  MSD-Value    |            Padding            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 9: P-flag in SRv6-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and can be removed at the time of publication.

## B.2.  SRv6 PCE Capability Flags

SRv6 PCE Capability TLV is defined in defined in [I-D.ietf-pce-segment-routing-ipv6], and the registry to manage the Flag field of the SRv6 PCE Capability Flags is requested in [I-D.ietf-pce-segment-routing-ipv6].  IANA is requested to make the following allocation in the aforementioned registry.

| Bit | Description | Reference |
|---|---|---|
| TBA2 | Path Segment Allocation is supported(P) | This document |

### B.3. Path Segment TLV

A new assignment should be done to the "PATH-SEGMENT TLV Segment Type (ST) Field" sub-registry for SRv6.

| Value | Description | Reference |
|---|---|---|
| 1 | SRv6 Path Segment (IPv6 addr) | This document |
| 2-255 | Reserved for future use | This document |

## Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: chengli13@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: Mach.chen@huawei.com

Weiqiang Cheng
China Mobile
China

Email: chengweiqiang@chinamobile.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: jie.dong@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing  100095
China

Email: lizhenbin@huawei.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

Quan Xiong
ZTE Corporation
China

Email: xiong.quan@zte.com.cn

PCE Working Group M. Negi
Internet-Draft D. Dhody
Intended status: Standards Track Huawei Technologies
Expires: April 22, 2019 S. Sivabalan
Cisco Systems
P. Kaladharan
RtBrick Inc
October 19, 2018

# PCEP Extensions for Segment Routing leveraging the IPv6 data plane
draft-negi-pce-segment-routing-ipv6-03

## Abstract

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. Segment Routing (SR) enables any head-end node to select any path without relying on a hop-by-hop signaling technique (e.g., LDP or RSVP-TE).

It depends only on "segments" that are advertised by Link- State IGPs. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Since, Segment Routing can be applied to both MPLS and IPv6 forwarding plane, a PCE should be able to compute SR-Path for both MPLS and IPv6 forwarding plane. This draft describes the extensions required for Segment Routing support for IPv6 data plane in Path Computation Element communication Protocol (PCEP).

## Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## Status of This Memo

Copyright Notice



Table of Contents

## 1. Introduction

As per [RFC8402], with Segment Routing (SR), a node steers a packet through an ordered list of instructions, called segments. A segment can represent any instruction, topological or service-based. A segment can have a semantic local to an SR node or global within an SR domain. SR allows to enforce a flow through any path and service chain while maintaining per-flow state only at the ingress node of the SR domain. Segments can be derived from different components: IGP, BGP, Services, Contexts, Locater, etc. The list of segment forming the path is called the Segment List and is encoded in the packet header. Segment Routing can be applied to the IPv6 architecture with the Segment Routing Header (SRH) [I-D.ietf-6man-segment-routing-header]. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment.

Segment Routing use cases are described in [RFC7855] and [RFC8354]. Segment Routing protocol extensions are defined in [I-D.ietf-isis-segment-routing-extensions], and [I-D.ietf-ospf-ospfv3-segment-routing-extensions].

As per [I-D.ietf-6man-segment-routing-header], an SRv6 Segment is a 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment". Further details are in an illustration provided in [I-D.filsfils-spring-srv6-network-programming].

The SR architecture can be applied to the MPLS forwarding plane without any change, in which case an SR path corresponds to an MPLS Label Switching Path (LSP). The SR is applied to IPV6 forwarding plane using SRH. A SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool, or a PCE and provisioned on the ingress node.

[RFC5440] describes Path Computation Element communication Protocol (PCEP) for communication between a Path Computation Client (PCC) and a Path Computation Element (PCE) or between a pair of PCEs. A PCE or a PCC operating as a PCE (in hierarchical PCE environment) computes paths for MPLS Traffic Engineering LSPs (MPLS-TE LSPs) based on various constraints and optimization criteria. [RFC8231] specifies extensions to PCEP that allow a stateful PCE to compute and recommend network paths in compliance with [RFC4657] and defines objects and TLVs for MPLS-TE LSPs. Stateful PCEP extensions provide synchronization of LSP state between a PCC and a PCE or between a pair of PCEs, delegation of LSP control, reporting of LSP state from a PCC to a PCE, controlling the setup and path routing of an LSP from a PCE to a PCC. Stateful PCEP extensions are intended for an operational model in which LSPs are configured on the PCC, and control over them is delegated to the PCE.

A mechanism to dynamically initiate LSPs on a PCC based on the requests from a stateful PCE or a controller using stateful PCE is specified in [RFC8281]. As per [I-D.ietf-pce-segment-routing], it is possible to use a stateful PCE for computing one or more SR-TE paths taking into account various constraints and objective functions. Once a path is chosen, the stateful PCE can initiate an SR-TE path on a PCC using PCEP extensions specified in [RFC8281] using the SR specific PCEP extensions specified in [I-D.ietf-pce-segment-routing]. [I-D.ietf-pce-segment-routing] specifies PCEP extensions for supporting a SR-TE LSP for MPLS data plane. This document extends [I-D.ietf-pce-segment-routing] to support SR for IPv6 data plane. Additionally, using procedures described in this document, a PCC can request an SRv6 path from either stateful or a stateless PCE. This specification relies on the PATH-SETUP-TYPE TLV and procedures specified in [RFC8408].

This specification provides a mechanism for a network controller (acting as a PCE) to instantiate candidate paths for an SR Policy onto a head-end node (acting as a PCC) using PCEP. For more information on the SR Policy Architecture, see [I-D.ietf-spring-segment-routing-policy].

## 2. Terminology

This document uses the following terms defined in [RFC5440]: PCC, PCE, PCEP Peer.

This document uses the following terms defined in [RFC8051]: Stateful PCE, Delegation.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

NAI: Node or Adjacency Identifier.

PCC: Path Computation Client.

PCE: Path Computation Element.

PCEP: Path Computation Element Protocol.

SR: Segment Routing.

SID: Segment Identifier.

SRv6: Segment Routing for IPv6 forwarding plane.

SRH: IPv6 Segment Routing Header.

SR Path: IPv6 Segment (List of IPv6 SID representing a path in IPv6 SR domain)

## 3. Overview of PCEP Operation in SRv6 Networks

Basic operations for PCEP speakers is as per [I-D.ietf-pce-segment-routing]. SRv6 Paths computed by a PCE can be represented as an ordered list of SRv6 segments of 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment" in this document.

[I-D.ietf-pce-segment-routing] defined a new ERO subobject denoted by "SR-ERO subobject" capable of carrying a SID as well as the identity of the node/adjacency represented by the SID. SR-capable PCEP speakers should be able to generate and/or process such ERO subobject. An ERO containing SR-ERO subobjects can be included in the PCEP Path Computation Reply (PCRep) message defined in [RFC5440], the PCEP LSP Initiate Request message (PCInitiate) defined in [RFC8281], as well as in the PCEP LSP Update Request (PCUpd) and PCEP LSP State Report (PCRpt) messages defined in defined in [RFC8231].

This document extends the "SR-ERO subobject" defined in [I-D.ietf-pce-segment-routing] to carry IPv6 SID(s) (IPv6 Addresses). SRv6-capable PCEP speakers MUST be able to generate and/or process this.

When a PCEP session between a PCC and a PCE is established, both PCEP speakers exchange their capabilities to indicate their ability to support SRv6 specific functionality.

In summary, this document:

o Defines a new PCEP capability for SRv6

- o Update the SR-ERO and SR-RRO sub-object for SRv6

- o Defines a new path setup type carried in the PATH-SETUP-TYPE and PATH-SETUP-TYPE-CAPABILITY TLVs.

## 3.1. Operation Overview

In SR networks, an ingress node of an SR path appends all outgoing packets with an SR header consisting of a list of SIDs (IPv6 Prefix in case of SRv6). The header has all necessary information to guide the packets from the ingress node to the egress node of the path, and hence there is no need for any signaling protocol.

For IPv6 in control plane with MPLS data-plane, mechanism remains same as [I-D.ietf-pce-segment-routing]

This document describes extensions to SR path for IPv6 data plane. SRv6 Path (i.e. ERO) consists of an ordered set of SRv6 SIDs(see details in Figure 2).

A PCC or PCE indicates its ability to support SRv6 during the PCEP session Initialization Phase via a new SRv6-PCE-CAPABILITY sub-TLV (see details in Section 3.3.1.1).

## 3.2. SRv6-Specific PCEP Message Extensions

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable length body made up of mandatory and/or optional objects. This document does not require any changes in the format of PCReq and PCRep messages specified in [RFC5440], PCInitiate message specified in [RFC8281], and PCRpt and PCUpd messages specified in [RFC8231]. However, PCEP messages pertaining to SRv6 MUST include PATH-SETUP-TYPE TLV in the RP or SRP object to clearly identify that SRv6 is intended.

## 3.3. Object Formats

### 3.3.1. The OPEN Object

#### 3.3.1.1. The SRv6 PCE Capability sub-TLV

This document defines a new Path Setup Type (PST) [RFC8408] for SRv6, as follows:

- o PST = TBD2: Path is setup using SRv6.

A PCEP speaker SHOULD indicate its support of the function described in this document by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this new PST included in the PST list.

This document also defines the SRv6-PCE-CAPABILITY sub-TLV. PCEP speakers use this sub-TLV to exchange information about their SRv6 capability. If a PCEP speaker includes PST=TBD2 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV then it MUST also include the SRv6-PCE- CAPABILITY sub-TLV inside the PATH-SETUP-TYPE-CAPABILITY TLV.

The format of the SRv6-PCE-CAPABILITY sub-TLV is shown in the following figure:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Type=TBD1          |            Length=4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      max-SL   |   Reserved    |             Flags           |L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: SRv6-PCE-CAPABILITY sub-TLV format

The code point for the TLV type (TBD1) is to be defined by IANA. The TLV length is 4 octets.

The 4-octet value comprise of -

max-SL: 1 octet, this field specifies the maximum value of the "Segments Left (SL)" in the SRH [I-D.ietf-6man-segment-routing-header].

Reserved: 1 octet, this field MUST be set to 0 on transmission, and ignored on receipt.

Flags: 2 octet, one bit is currently assigned in this document.

L bit: A PCC sets this bit to 1 to indicate that it does not impose any limit on SL.

Unassigned bits MUST be set to 0 and ignored on receipt.

#### 3.3.1.2. Exchanging the SRv6 Capability

A PCC indicates that it is capable of supporting the head-end functions for SRv6 by including the SRv6-PCE-CAPABILITY sub-TLV in the Open message that it sends to a PCE. A PCE indicates that it is capable of computing SRv6 paths by including the SRv6-PCE-CAPABILITY sub-TLV in the Open message that it sends to a PCC.

If a PCEP speaker receives a PATH-SETUP-TYPE-CAPABILITY TLV with a PST list containing PST=TBD2, but the SRv6-PCE-CAPABILITY sub-TLV is absent, then the PCEP speaker MUST send a PCErr message with Error-Type 10 (Reception of an invalid object) and Error-Value TBD5 (to be assigned by IANA) (Missing PCE-SRv6-CAPABILITY sub-TLV) and MUST then close the PCEP session. If a PCEP speaker receives a PATH-SETUP-TYPE-CAPABILITY TLV with a SRv6-PCE-CAPABILITY sub-TLV, but the PST list does not contain PST=TBD2, then the PCEP speaker MUST ignore the SRv6-PCE-CAPABILITY sub-TLV.

The number of SRv6 SIDs that can be imposed on a packet depends on the PCC's IPv6 data plane's capability. If a PCC sets the L flag to 1 then the max-SL is not used and MUST be set to zero. If a PCE receives an SRv6-PCE-CAPABILITY sub-TLV with the L flag set to 1 then it MUST ignore the max-SL field and MUST assume that the sender can impose a SL of any length. If a PCC sets the L flag to zero, then it sets the max-SL field to the maximum number of SIDs that it can impose on a packet. If a PCE receives an SRv6-PCE-CAPABILITY sub-TLV with the L flag and max-SL both set to zero then it MUST assume that the PCC is not capable of imposing a SL of any length and hence is not SRv6 capable, unless it learns a non-zero max-SL for the PCC through some other means.

Note that the max-SL value exchanged via the SRv6-PCE-CAPABILITY sub-TLV indicates the SRv6 SID imposition limit for the PCC node. However, if a PCE learns the max-SL value of a PCC node via different means, e.g routing protocols, as specified in: [I-D.li-ospf-ospfv3-srv6-extensions]; [I-D.bashandy-isis-srv6-extensions]; [I-D.dawra-idr-bgpls-srv6-ext], then it ignores the max-SL value in the SRv6-PCE-CAPABILITY sub-TLV. Furthermore, whenever a PCE learns the other advanced max-SL via different means, it MUST use that value regardless of the max-SL value exchanged in the SRv6-PCE-CAPABILITY sub-TLV.

Once an SRv6-capable PCEP session is established with a non-zero max-SL value, the corresponding PCE MUST NOT send SRv6 paths with a number of SIDs exceeding that max-SL value. If a PCC needs to modify the max-SL value, it MUST close the PCEP session and re-establish it with the new max-SL value. If a PCEP session is established with a non-zero max-SL value, and the PCC receives an SRv6 path containing more SIDs than specified in the max-SL value, the PCC MUST send a PCErr message with Error-Type 10 (Reception of an invalid object) and Error-Value 3 (Unsupported number of Segment ERO subobjects). If a PCEP session is established with an max-SL value of zero, then the PCC MAY specify an max-SL for each path computation request that it sends to the PCE, by including a "maximum SID depth" metric object on the request similar to [I-D.ietf-pce-segment-routing].

The L flag and Max-SL value inside the SRv6-PCE-CAPABILITY sub-TLV are meaningful only in the Open message sent from a PCC to a PCE. As such, a PCE MUST set the L flag and Max-SL value to zero in an outbound message to a PCC. Similarly, a PCC MUST ignore any max-SL value received from a PCE. If a PCE receives multiple SRv6-PCE-CAPABILITY sub-TLVs in an Open message, it processes only the first sub-TLV received.

#### 3.3.2. The RP/SRP Object

In order to indicate the SRv6 path, RP or SRP object MUST include the PATH-SETUP-TYPE TLV specified in [RFC8408]. This document defines a new Path Setup Type (PST=TBD2) for SRv6.

The LSP-IDENTIFIERS TLV MAY be present for the above PST type.

#### 3.3.3. ERO Object

In order to support SRv6, the SR-ERO subobject is used [I-D.ietf-pce-segment-routing]. This documents extends the SR-ERO subobject. All the processing rules remains the same.

##### 3.3.3.1. SR-ERO Subobject

For supporting SRv6, a new NAI Type (NT) is defined, the format of SR-ERO sub object remains the same as defined in [I-D.ietf-pce-segment-routing].

When the NAI Type (NT) indicates SRv6, then the SR-ERO subobject represent a SRv6 segment and include a field SRv6I (SRv6 Identifier) in place of NAI (Node or Adjacency Identifier) defined in [I-D.ietf-pce-segment-routing]. The 32 bit SID is not used for SRv6 and MUST NOT be included. The format of SR-ERO subobject is reproduced with the SRv6I field as shown below:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |L|    Type     |     Length    |  NT   |     Flags     |F|S|C|M|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     SID (not included)                        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //                      SRv6I (variable)                       //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: SR-ERO Subobject Format

The description of all the flags and fields is as per [I-D.ietf-pce-segment-routing].

For SRv6 segments, a new NT (NAI Type) is assigned by IANA as TBD3.

For SRv6 segments (when NT is TBD3), M and C flag MUST NOT be set. The S flag MUST be set and SID field MUST NOT be included. The F bit MUST NOT be set.

If these flags are not set properly, the subobject MUST be considered malformed and the PCEP speaker react as per the error handling described in Section 3.3.3.2.

The SRv6I format is as shown below:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    |                       SRv6 Identifier                         |
    |                          (128-bit)                            |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | SRv6NT|         Flags         |        Function Code          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //                         NAI (variable)                      //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: SR-ERO Subobject's SRv6I Format

SRv6 Identifier is the 128 bit IPv6 addresses representing SRv6 segment.

SRv6NT is the SRv6 NAI Type which indicates the interpretation for NAI (Node or Adjacency Identifier) as per [I-D.ietf-pce-segment-routing].

Flags is the 12 bit field, no flag bits are currently defined in this document. This MUST be set to 0 and ignored on receipt.

Function Code is is the 16 bit field representing supported functions associated with SRv6 SIDs. This information is optional and included only for maintainability. Following function codes are currently defined -

0: Reserved

1: End Function

2: End.DX6 Function

3: End.DT6 Function

4: End.X Function

NAI field [I-D.ietf-pce-segment-routing] contains the NAI associated with the SRv6 Identifier. Depending on the value of SRv6NT, the NAI can have different formats.

When SRv6NT value is 1, the NAI is as per the 'IPv6 Node ID' format defined in [I-D.ietf-pce-segment-routing], which specify an IPv6 address. This is used to identify the owner of the SRv6 Identifier. This is optional, as the LOC (the locater portion) of the SRv6 SID serves a similar purpose.

When SRv6NT value is 2, the NAI is as per the 'IPv6 Adjacency' format defined in [I-D.ietf-pce-segment-routing], which specify a pair of IPv6 addresses. This is used to identify the IPv6 Adjacency and used with the SRv6 Adj-SID.

Note that when SRv6NT value is 0, NAI is not included and MUST be NULL.

#### 3.3.3.2. ERO Processing

The ERO and SR-ERO subobject processing remains as per [RFC5440] and [I-D.ietf-pce-segment-routing].

The NT MUST only be TDB3, if the PST=TBD3 is set in the PCEP message and SRv6-PCE-CAPABILITY sub-TLV is exchanged with the PCEP peer. In case a PCEP speaker receives the SR-ERO subobject with NT indicating

SRv6 segment, when the PST is not set to TBD3 or SRv6-PCE-CAPABILITY sub-TLV was not exchanged, it MUST send a PCErr message with Error-Type = 19 ("Invalid Operation") and Error-Value = TBD4 ("Attempted SRv6 when the capability was not advertised"). A PCEP speaker that does not recognize the NT value, it would behave as per [I-D.ietf-pce-segment-routing].

If a PCC receives a list of SRv6 segments, and the number of SRv6 segments exceeds the max-SL that the PCC can impose on the packet (SRH), it MAY send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Unsupported number of Segment ERO subobjects") as per [I-D.ietf-pce-segment-routing].

When a PCEP speaker detects that all subobjects of ERO are not identical to SRv6, and if it does not handle such ERO, it MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Non-identical ERO subobjects")as per [I-D.ietf-pce-segment-routing].

When a PCEP speaker receives an SR-ERO subobject for SRv6 segment, M, C and F flag MUST NOT be set and S flag MUST be set. Otherwise, it MUST consider the entire ERO object invalid and send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Malformed object") as per [I-D.ietf-pce-segment-routing]. The PCEP speaker MAY include the malformed SR-ERO object in the PCErr message as well.

###### 3.3.3.2.1. Interpreting the SR-ERO

The SR-ERO contains a sequence of subobjects. According to [I-D.ietf-spring-segment-routing-policy], each SR-ERO subobject in the sequence identifies a segment that the traffic will be directed to, in the order given. That is, the first subobject identifies the first segment the traffic will be directed to, the second SR-ERO subobject represents the second segment, and so on.

The PCC interprets the SR-ERO by converting it to an SRv6 SRH plus a next hop. The PCC sends packets along the segment routed path by prepending the SRH onto the packets and sending the resulting, modified packet to the next hop.

#### 3.3.4. RRO Object

In order to support SRv6, the SR-RRO Subobject is used [I-D.ietf-pce-segment-routing]. All other processing rules remains the same.

#### 3.3.4.1. SR-RRO Subobject

For SRv6 segments, a new NT (NAI Type) is assigned by IANA as TBD3, the format of SR-RRO sub object remains the same as the SR-ERO subobject, but without the L flag [I-D.ietf-pce-segment-routing].

When the NAI Type (NT) indicates SRv6, then the SR-RRO subobject represent a SRv6 segment and include a field SRv6I (SRv6 Identifier) in place of NAI (Node or Adjacency Identifier) defined in [I-D.ietf-pce-segment-routing]. The 32 bit SID MUST NOT be included. The format of SR-RRO subobject is reproduced with the SRv6I field as shown below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |     Length    |  ST   |     Flags     |F|S|C|M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     SID (not included)                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                     SRv6I (variable)                        //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: SR-RRO Subobject Format

The description of all fields and flags is as per SR-ERO subobject.

Processing rules of SR-RRO subobject are identical to those of SR-ERO subobject.

If a PCE detects that all subobjects of RRO are not identical, and if it does not handle such RRO, it MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = 10 ("Non-identical RRO subobjects").

### 3.4. Security Considerations

The security considerations described in [RFC5440], [RFC8231] and [RFC8281], [I-D.ietf-pce-segment-routing], are applicable to this specification. No additional security measure is required.

### 3.5. IANA Considerations

This document requests IANA to include (I) bit in flags registry for SR-ERO and SR-RRO sub-objects. Other changes are defined as:

#### 3.5.1. PCEP Objects

##### 3.5.1.1. ERROR Objects

IANA is requested to allocate code-points in the PCEP-ERROR Object Error Types and Values registry for the following new error-values:

| Error-Type | Meaning |
|---|---|
| 10 | Reception of an invalid object<br>Error-value = TBD5 (Missing PCE-SRv6-CAPABILITY sub-TLV) |
| 19 | Invalid Operation<br>Error-value = TBD4 (Attempted SRv6 when the capability was not advertised) |

##### 3.5.1.2. TLV Type Indicators

IANA is requested to make the assignment of the new code points for the existing "PCEP TLV Type Indicators" registry as follows:

| Value | Meaning | Reference |
|---|---|---|
| TBD1 | SRv6-PCE-CAPABILITY | This Document |

##### 3.5.1.3. New Path Setup Type

[RFC8408] 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 |
|---|---|---|
| TBD2 | SRv6 (SRH) technique | This Document |

### 3.6. The NAI Type field

As per [I-D.ietf-pce-segment-routing], a new subregistery for "PCEP SR-ERO NAI Types" was created. IANA is requested to make the assignment of the new code points in the afore-mentioned registry as follows:

| Value | Description | Reference |
| --- | --- | --- |
| TBD3 | NAI is an SRv6 segment | This Document |

# 4. References

## 4.1. 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>.

[RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <https://www.rfc-editor.org/info/rfc5511>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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>.

[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-14 (work in progress), October 2018.

## 4.2. Informative References

[RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <https://www.rfc-editor.org/info/rfc4657>.

[RFC7855] Previdi, S., Ed., Filsfils, C., Ed., Decraene, B., Litkowski, S., Horneffer, M., and R. Shakir, "Source Packet Routing in Networking (SPRING) Problem Statement and Requirements", RFC 7855, DOI 10.17487/RFC7855, May 2016, <https://www.rfc-editor.org/info/rfc7855>.

[RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <https://www.rfc-editor.org/info/rfc8051>.

[RFC8354] Brzozowski, J., Leddy, J., Filsfils, C., Maglione, R., Ed., and M. Townsley, "Use Cases for IPv6 Source Packet Routing in Networking (SPRING)", RFC 8354, DOI 10.17487/RFC8354, March 2018, <https://www.rfc-editor.org/info/rfc8354>.

[I-D.ietf-6man-segment-routing-header]
Filsfils, C., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-14 (work in progress), June 2018.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., Decraene, B., and J. Tantsura, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-19 (work in progress), July 2018.

[I-D.ietf-ospf-ospfv3-segment-routing-extensions]
Psenak, P., Filsfils, C., Previdi, S., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPFv3 Extensions for Segment Routing", draft-ietf-ospf-ospfv3-segment-routing-extensions-15 (work in progress), August 2018.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d., bogdanov@google.com, b., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-01 (work in progress), June 2018.

[I-D.filsfils-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-filsfils-spring-srv6-network-programming-05 (work in progress), July 2018.

[I-D.li-ospf-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", draft-li-ospf-ospfv3-srv6-extensions-02 (work in progress), September 2018.

[I-D.bashandy-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Routing over IPv6 Dataplane", draft-bashandy-isis-srv6-extensions-04 (work in progress), October 2018.

[I-D.dawra-idr-bgpls-srv6-ext]
Dawra, G., Filsfils, C., Talaulikar, K., Chen, M., daniel.bernier@bell.ca, d., Uttaro, J., Decraene, B., and H. Elmalky, "BGP Link State extensions for IPv6 Segment Routing(SRv6)", draft-dawra-idr-bgpls-srv6-ext-04 (work in progress), September 2018.

## Appendix A. Contributor

The following persons contributed to this document:

Huang Wumin
Huawei Technologies
Huawei Building, No. 156 Beiqing Rd.
Beijing 100095
China

Email: huangwumin@huawei.com

## Authors' Addresses

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahendrasingh@huawei.com

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Siva Sivabalan
Cisco Systems

EMail: msiva@cisco.com

Prejeeth Kaladharan
RtBrick Inc
Bangalore, Karnataka
India

EMail: prejeeth@rtbrick.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 10, 2019

M. Negi
C. Li
Huawei Technologies
S. Sivabalan
Cisco Systems
P. Kaladharan
RtBrick Inc

February 6, 2019

# PCEP Extensions for Segment Routing leveraging the IPv6 data plane
draft-negi-pce-segment-routing-ipv6-04

## Abstract

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. Segment Routing (SR) enables any head-end node to select any path without relying on a hop-by-hop signaling technique (e.g., LDP or RSVP-TE).

It depends only on "segments" that are advertised by Link- State IGPs. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Since, Segment Routing can be applied to both MPLS and IPv6 forwarding plane, a PCE should be able to compute SR-Path for both MPLS and IPv6 forwarding plane. This draft describes the extensions required for Segment Routing support for IPv6 data plane in Path Computation Element communication Protocol (PCEP).

## Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## Status of This Memo

Copyright Notice



Table of Contents

## 1. Introduction

As per [RFC8402], with Segment Routing (SR), a node steers a packet through an ordered list of instructions, called segments. A segment can represent any instruction, topological or service-based. A segment can have a semantic local to an SR node or global within an SR domain. SR allows to enforce a flow through any path and service chain while maintaining per-flow state only at the ingress node of the SR domain. Segments can be derived from different components: IGP, BGP, Services, Contexts, Locater, etc. The list of segment forming the path is called the Segment List and is encoded in the packet header. Segment Routing can be applied to the IPv6 architecture with the Segment Routing Header (SRH) [I-D.ietf-6man-segment-routing-header]. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment.

Segment Routing use cases are described in [RFC7855] and [RFC8354]. Segment Routing protocol extensions are defined in [I-D.ietf-isis-segment-routing-extensions], and [I-D.ietf-ospf-ospfv3-segment-routing-extensions].

As per [I-D.ietf-6man-segment-routing-header], an SRv6 Segment is a 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment". Further details are in an illustration provided in [I-D.filsfils-spring-srv6-network-programming].

The SR architecture can be applied to the MPLS forwarding plane without any change, in which case an SR path corresponds to an MPLS Label Switching Path (LSP). The SR is applied to IPV6 forwarding plane using SRH. A SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool, or a PCE and provisioned on the ingress node.

[RFC5440] describes Path Computation Element communication Protocol (PCEP) for communication between a Path Computation Client (PCC) and a Path Computation Element (PCE) or between a pair of PCEs. A PCE or a PCC operating as a PCE (in hierarchical PCE environment) computes paths for MPLS Traffic Engineering LSPs (MPLS-TE LSPs) based on various constraints and optimization criteria. [RFC8231] specifies extensions to PCEP that allow a stateful PCE to compute and recommend network paths in compliance with [RFC4657] and defines objects and TLVs for MPLS-TE LSPs. Stateful PCEP extensions provide synchronization of LSP state between a PCC and a PCE or between a pair of PCEs, delegation of LSP control, reporting of LSP state from a PCC to a PCE, controlling the setup and path routing of an LSP from a PCE to a PCC. Stateful PCEP extensions are intended for an operational model in which LSPs are configured on the PCC, and control over them is delegated to the PCE.

A mechanism to dynamically initiate LSPs on a PCC based on the requests from a stateful PCE or a controller using stateful PCE is specified in [RFC8281]. As per [I-D.ietf-pce-segment-routing], it is possible to use a stateful PCE for computing one or more SR-TE paths taking into account various constraints and objective functions. Once a path is chosen, the stateful PCE can initiate an SR-TE path on a PCC using PCEP extensions specified in [RFC8281] using the SR specific PCEP extensions specified in [I-D.ietf-pce-segment-routing]. [I-D.ietf-pce-segment-routing] specifies PCEP extensions for supporting a SR-TE LSP for MPLS data plane. This document extends [I-D.ietf-pce-segment-routing] to support SR for IPv6 data plane. Additionally, using procedures described in this document, a PCC can request an SRv6 path from either stateful or a stateless PCE. This specification relies on the PATH-SETUP-TYPE TLV and procedures specified in [RFC8408].

This specification provides a mechanism for a network controller (acting as a PCE) to instantiate candidate paths for an SR Policy onto a head-end node (acting as a PCC) using PCEP. For more information on the SR Policy Architecture, see [I-D.ietf-spring-segment-routing-policy].

## 2. Terminology

This document uses the following terms defined in [RFC5440]: PCC, PCE, PCEP Peer.

This document uses the following terms defined in [RFC8051]: Stateful PCE, Delegation.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

NAI: Node or Adjacency Identifier.

PCC: Path Computation Client.

PCE: Path Computation Element.

PCEP: Path Computation Element Protocol.

SR: Segment Routing.

SID: Segment Identifier.

SRv6: Segment Routing for IPv6 forwarding plane.

SRH: IPv6 Segment Routing Header.

SR Path: IPv6 Segment (List of IPv6 SID representing a path in IPv6 SR domain)

## 3. Overview of PCEP Operation in SRv6 Networks

Basic operations for PCEP speakers is as per [I-D.ietf-pce-segment-routing]. SRv6 Paths computed by a PCE can be represented as an ordered list of SRv6 segments of 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment" in this document.

[I-D.ietf-pce-segment-routing] defined a new ERO subobject denoted by "SR-ERO subobject" capable of carrying a SID as well as the identity of the node/adjacency represented by the SID. SR-capable PCEP speakers should be able to generate and/or process such ERO subobject. An ERO containing SR-ERO subobjects can be included in the PCEP Path Computation Reply (PCRep) message defined in [RFC5440], the PCEP LSP Initiate Request message (PCInitiate) defined in [RFC8281], as well as in the PCEP LSP Update Request (PCUpd) and PCEP LSP State Report (PCRpt) messages defined in defined in [RFC8231].

This document extends the "SR-ERO subobject" defined in [I-D.ietf-pce-segment-routing] to carry IPv6 SID(s) (IPv6 Addresses). SRv6-capable PCEP speakers MUST be able to generate and/or process this.

When a PCEP session between a PCC and a PCE is established, both PCEP speakers exchange their capabilities to indicate their ability to support SRv6 specific functionality.

In summary, this document:

- o Defines a new PCEP capability for SRv6
- o Update the SR-ERO and SR-RRO sub-object for SRv6
- o Defines a new path setup type carried in the PATH-SETUP-TYPE and PATH-SETUP-TYPE-CAPABILITY TLVs.

## 3.1. Operation Overview

In SR networks, an ingress node of an SR path appends all outgoing packets with an SR header consisting of a list of SIDs (IPv6 Prefix in case of SRv6). The header has all necessary information to guide the packets from the ingress node to the egress node of the path, and hence there is no need for any signaling protocol.

For IPv6 in control plane with MPLS data-plane, mechanism remains same as [I-D.ietf-pce-segment-routing]

This document describes extensions to SR path for IPv6 data plane. SRv6 Path (i.e. ERO) consists of an ordered set of SRv6 SIDs(see details in Figure 2).

A PCC or PCE indicates its ability to support SRv6 during the PCEP session Initialization Phase via a new SRv6-PCE-CAPABILITY sub-TLV (see details in Section 3.3.1.1).

## 3.2. SRv6-Specific PCEP Message Extensions

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable length body made up of mandatory and/or optional objects. This document does not require any changes in the format of PCReq and PCRep messages specified in [RFC5440], PCInitiate message specified in [RFC8281], and PCRpt and PCUpd messages specified in [RFC8231]. However, PCEP messages pertaining to SRv6 MUST include PATH-SETUP-TYPE TLV in the RP or SRP object to clearly identify that SRv6 is intended.

## 3.3. Object Formats

### 3.3.1. The OPEN Object

#### 3.3.1.1. The SRv6 PCE Capability sub-TLV

This document defines a new Path Setup Type (PST) [RFC8408] for SRv6, as follows:

- o PST = TBD2: Path is setup using SRv6.

A PCEP speaker SHOULD indicate its support of the function described in this document by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this new PST included in the PST list.

This document also defines the SRv6-PCE-CAPABILITY sub-TLV. PCEP speakers use this sub-TLV to exchange information about their SRv6 capability. If a PCEP speaker includes PST=TBD2 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV then it MUST also include the SRv6-PCE- CAPABILITY sub-TLV inside the PATH-SETUP-TYPE-CAPABILITY TLV.

The format of the SRv6-PCE-CAPABILITY sub-TLV is shown in the following figure:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Type=TBD1          |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Reserved           |             Flags           |L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   MSD-Type    |  MSD-Value    |   MSD-Type    |  MSD-Value    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                              ...                            //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   MSD-Type    |  MSD-Value    |            Padding            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: SRv6-PCE-CAPABILITY sub-TLV format

The code point for the TLV type (TBD1) is to be defined by IANA. The TLV length is variable.

The value comprise of -

Reserved: 2 octet, this field MUST be set to 0 on transmission, and ignored on receipt.

Flags: 2 octet, one bit is currently assigned in this document.

L bit: A PCC sets this bit to 1 to indicate that it does not impose any limit on MSD (irrespective of the MSD-Type).

Unassigned bits MUST be set to 0 and ignored on receipt.

A pair of (MSD-Type,MSD-Value): Where MSD-Type (1 octet) is as per the IGP MSD Type registry created by [RFC8491] and populated with SRv6 MSD types as per [I-D.bashandy-isis-srv6-extensions]; MSD-Value (1 octet) is as per [RFC8491].

The TLV format is compliant with the PCEP TLV format defined in [RFC5440]. That is, the TLV is composed of 2 octets for the type, 2 octets specifying the TLV length, and a Value field. The Length field defines the length of the value portion in octets. The TLV is padded to 4-octet alignment, and padding is not included in the Length field. The number of (MSD-Type,MSD-Value) pairs can be determined from the Length field of the TLV.

#### 3.3.1.2. Exchanging the SRv6 Capability

A PCC indicates that it is capable of supporting the head-end functions for SRv6 by including the SRv6-PCE-CAPABILITY sub-TLV in the Open message that it sends to a PCE. A PCE indicates that it is capable of computing SRv6 paths by including the SRv6-PCE-CAPABILITY sub-TLV in the Open message that it sends to a PCC.

If a PCEP speaker receives a PATH-SETUP-TYPE-CAPABILITY TLV with a PST list containing PST=TBD2, but the SRv6-PCE-CAPABILITY sub-TLV is absent, then the PCEP speaker MUST send a PCErr message with Error-Type 10 (Reception of an invalid object) and Error-Value TBD5 (to be assigned by IANA) (Missing PCE-SRv6-CAPABILITY sub-TLV) and MUST then close the PCEP session. If a PCEP speaker receives a PATH-SETUP-TYPE-CAPABILITY TLV with a SRv6-PCE-CAPABILITY sub-TLV, but the PST list does not contain PST=TBD2, then the PCEP speaker MUST ignore the SRv6-PCE-CAPABILITY sub-TLV.

The number of SRv6 SIDs that can be imposed on a packet depends on the PCC's IPv6 data plane's capability. If a PCC sets the L flag to 1 then the MSD is not used and MUST not be included. If a PCE receives an SRv6-PCE-CAPABILITY sub-TLV with the L flag set to 1 then it MUST ignore any MSD-Type, MSD-Value fields and MUST assume that the sender can impose any length of SRH. If a PCC sets the L flag to zero, then it sets the SRv6 MSD-Type, MSD-Value fields that it can impose on a packet. If a PCE receives an SRv6-PCE-CAPABILITY sub-TLV with the L flag and SRv6 MSD-Type, MSD-Value fields both set to zero then it is considered as an error and the PCE MUST respond with a PCErr message (Error-Type=1 "PCEP session establishment failure" and Error-Value=1 "reception of an invalid Open message or a non Open message."). In case the MSD-Type in SRv6-PCE-CAPABILITY sub-TLV received by the PCE does not correspond to one of the SRv6 MSD types, the PCE MUST respond with a PCErr message (Error-Type=1 "PCEP session establishment failure" and Error-Value=1 "reception of an invalid Open message or a non Open message.").

Note that the MSD-Type, MSD-Value exchanged via the SRv6-PCE-CAPABILITY sub-TLV indicates the SRv6 SID imposition limit for the PCC node. However, if a PCE learns these via different means, e.g routing protocols, as specified in: [I-D.li-ospf-ospfv3-srv6-extensions]; [I-D.bashandy-isis-srv6-extensions]; [I-D.dawra-idr-bgpls-srv6-ext], then it ignores the values in the SRv6-PCE-CAPABILITY sub-TLV. Furthermore, whenever a PCE learns the other advanced SRv6 MSD via different means, it MUST use that value regardless of the values exchanged in the SRv6-PCE-CAPABILITY sub-TLV.

Once an SRv6-capable PCEP session is established with a non-zero SRv6 MSD value, the corresponding PCE MUST NOT send SRv6 paths with a number of SIDs exceeding that SRv6 MSD value (based on the SRv6 MSD Type). If a PCC needs to modify the SRv6 MSD value, it MUST close the PCEP session and re-establish it with the new value. If a PCEP session is established with a non-zero SRv6 MSD value, and the PCC receives an SRv6 path containing more SIDs than specified in the SRv6 MSD value (based on the SRv6 MSD type), the PCC MUST send a PCErr message with Error-Type 10 (Reception of an invalid object) and Error-Value 3 (Unsupported number of Segment ERO subobjects). If a PCEP session is established with an SRv6 MSD value of zero, then the PCC MAY specify an SRv6 MSD for each path computation request that it sends to the PCE, by including a "maximum SID depth" metric object on the request similar to [I-D.ietf-pce-segment-routing].

The L flag and (MSD-Type,MSD-Value) pair inside the SRv6-PCE-CAPABILITY sub-TLV are meaningful only in the Open message sent from a PCC to a PCE. As such, a PCE MUST set the L flag and not include (MSD-Type,MSD-Value) in an outbound message to a PCC. Similarly, a PCC MUST ignore any (MSD-Type,MSD-Value) received from a PCE. If a PCE receives multiple SRv6-PCE-CAPABILITY sub-TLVs in an Open message, it processes only the first sub-TLV received.

#### 3.3.2. The RP/SRP Object

In order to indicate the SRv6 path, RP or SRP object MUST include the PATH-SETUP-TYPE TLV specified in [RFC8408]. This document defines a new Path Setup Type (PST=TBD2) for SRv6.

The LSP-IDENTIFIERS TLV MAY be present for the above PST type.

#### 3.3.3. ERO Object

In order to support SRv6, the SR-ERO subobject is used [I-D.ietf-pce-segment-routing]. This documents extends the SR-ERO subobject. All the processing rules remains the same.

### 3.3.3.1. SR-ERO Subobject

For supporting SRv6, a new NAI Type (NT) is defined, the format of SR-ERO sub object remains the same as defined in [I-D.ietf-pce-segment-routing].

When the NAI Type (NT) indicates SRv6, then the SR-ERO subobject represent a SRv6 segment and include a field SRv6I (SRv6 Identifier) in place of NAI (Node or Adjacency Identifier) defined in [I-D.ietf-pce-segment-routing]. The 32 bit SID is not used for SRv6 and MUST NOT be included. The format of SR-ERO subobject is reproduced with the SRv6I field as shown below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|L|    Type     |     Length    |  NT   |     Flags     |F|S|C|M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      SID (not included)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                      SRv6I (variable)                       //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: SR-ERO Subobject Format

The description of all the flags and fields is as per [I-D.ietf-pce-segment-routing].

For SRv6 segments, a new NT (NAI Type) is assigned by IANA as TBD3.

For SRv6 segments (when NT is TBD3), M and C flag MUST NOT be set. The S flag MUST be set and SID field MUST NOT be included. The F bit MUST NOT be set.

If these flags are not set properly, the subobject MUST be considered malformed and the PCEP speaker react as per the error handling described in Section 3.3.3.2.

The SRv6I format is as shown below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                      SRv6 Identifier                          |
|                         (128-bit)                             |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| SRv6NT|         Flags         |         Function Code         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                        NAI (variable)                       //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: SR-ERO Subobject's SRv6I Format

SRv6 Identifier is the 128 bit IPv6 addresses representing SRv6 segment.

SRv6NT is the SRv6 NAI Type which indicates the interpretation for NAI (Node or Adjacency Identifier) as per [I-D.ietf-pce-segment-routing].

Flags is the 12 bit field, no flag bits are currently defined in this document. This MUST be set to 0 and ignored on receipt.

Function Code is is the 16 bit field representing supported functions associated with SRv6 SIDs. This information is optional and included only for maintainability. Following function codes are currently defined -

0: Reserved

1: End Function

2: End.DX6 Function

3: End.DT6 Function

4: End.X Function

NAI field [I-D.ietf-pce-segment-routing] contains the NAI associated with the SRv6 Identifier. Depending on the value of SRv6NT, the NAI can have different formats.

When SRv6NT value is 1, the NAI is as per the 'IPv6 Node ID' format defined in [I-D.ietf-pce-segment-routing], which specify an IPv6 address. This is used to identify the owner of the

SRv6 Identifier. This is optional, as the LOC (the locater portion) of the SRv6 SID serves a similar purpose.

When SRv6NT value is 2, the NAI is as per the 'IPv6 Adjacency' format defined in [I-D.ietf-pce-segment-routing], which specify a pair of IPv6 addresses. This is used to identify the IPv6 Adjacency and used with the SRv6 Adj-SID.

Note that when SRv6NT value is 0, NAI is not included and MUST be NULL.

[Editor's Note - Add IPv6 unnumbered adjacency, once done by [I-D.ietf-pce-segment-routing]]

#### 3.3.3.2. ERO Processing

The ERO and SR-ERO subobject processing remains as per [RFC5440] and [I-D.ietf-pce-segment-routing].

The NT MUST only be TDB3, if the PST=TBD3 is set in the PCEP message and SRv6-PCE-CAPABILITY sub-TLV is exchanged with the PCEP peer. In case a PCEP speaker receives the SR-ERO subobject with NT indicating SRv6 segment, when the PST is not set to TBD3 or SRv6-PCE-CAPABILITY sub-TLV was not exchanged, it MUST send a PCErr message with Error-Type = 19 ("Invalid Operation") and Error-Value = TBD4 ("Attempted SRv6 when the capability was not advertised"). A PCEP speaker that does not recognize the NT value, it would behave as per [I-D.ietf-pce-segment-routing].

If a PCC receives a list of SRv6 segments, and the number of SRv6 segments exceeds the SRv6 MSD that the PCC can impose on the packet (SRH), it MAY send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Unsupported number of Segment ERO subobjects") as per [I-D.ietf-pce-segment-routing].

When a PCEP speaker detects that all subobjects of ERO are not identical to SRv6, and if it does not handle such ERO, it MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Non-identical ERO subobjects")as per [I-D.ietf-pce-segment-routing].

When a PCEP speaker receives an SR-ERO subobject for SRv6 segment, M, C and F flag MUST NOT be set and S flag MUST be set. Otherwise, it MUST consider the entire ERO object invalid and send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Malformed object") as per [I-D.ietf-pce-segment-routing]. The PCEP speaker MAY include the malformed SR-ERO object in the PCErr message as well.

#### 3.3.3.2.1. Interpreting the SR-ERO

The SR-ERO contains a sequence of subobjects. According to [I-D.ietf-spring-segment-routing-policy], each SR-ERO subobject in the sequence identifies a segment that the traffic will be directed to, in the order given. That is, the first subobject identifies the first segment the traffic will be directed to, the second SR-ERO subobject represents the second segment, and so on.

The PCC interprets the SR-ERO by converting it to an SRv6 SRH plus a next hop. The PCC sends packets along the segment routed path by prepending the SRH onto the packets and sending the resulting, modified packet to the next hop.

### 3.3.4. RRO Object

In order to support SRv6, the SR-RRO Subobject is used [I-D.ietf-pce-segment-routing]. All other processing rules remains the same.

#### 3.3.4.1. SR-RRO Subobject

For SRv6 segments, a new NT (NAI Type) is assigned by IANA as TBD3, the format of SR-RRO sub object remains the same as the SR-ERO subobject, but without the L flag [I-D.ietf-pce-segment-routing].

When the NAI Type (NT) indicates SRv6, then the SR-RRO subobject represent a SRv6 segment and include a field SRv6I (SRv6 Identifier) in place of NAI (Node or Adjacency Identifier) defined in [I-D.ietf-pce-segment-routing]. The 32 bit SID MUST NOT be included. The format of SR-RRO subobject is reproduced with the SRv6I field as shown below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Length    |  ST   |      Flags    |F|S|C|M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       SID (not included)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                       SRv6I (variable)                      //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: SR-RRO Subobject Format

The description of all fields and flags is as per SR-ERO subobject.

Processing rules of SR-RRO subobject are identical to those of SR-ERO subobject.

If a PCE detects that all subobjects of RRO are not identical, and if it does not handle such RRO, it MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = 10 ("Non-identical RRO subobjects").

## 3.4. Security Considerations

The security considerations described in [RFC5440], [RFC8231] and [RFC8281], [I-D.ietf-pce-segment-routing], are applicable to this specification. No additional security measure is required.

## 3.5. IANA Considerations

This document requests IANA to include (I) bit in flags registry for SR-ERO and SR-RRO sub-objects. Other changes are defined as:

### 3.5.1. PCEP Objects

#### 3.5.1.1. ERROR Objects

IANA is requested to allocate code-points in the PCEP-ERROR Object Error Types and Values registry for the following new error-values:

| Error-Type | Meaning |
|---|---|
| 10 | Reception of an invalid object<br>Error-value = TBD5 (Missing PCE-SRv6-CAPABILITY sub-TLV) |
| 19 | Invalid Operation<br>Error-value = TBD4 (Attempted SRv6 when the capability was not advertised) |

#### 3.5.1.2. TLV Type Indicators

IANA is requested to make the assignment of the new code points for the existing "PCEP TLV Type Indicators" registry as follows:

| Value | Meaning | Reference |
|---|---|---|
| TBD1 | SRv6-PCE-CAPABILITY | This Document |

#### 3.5.1.3. New Path Setup Type

[RFC8408] 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 |
| --- | --- | --- |
| TBD2 | SRv6 (SRH) technique | This Document |

### 3.6. The NAI Type field

As per [I-D.ietf-pce-segment-routing], a new subregistery for "PCEP SR-ERO NAI Types" was created. IANA is requested to make the assignment of the new code points in the afore-mentioned registry as follows:

| Value | Description | Reference |
| --- | --- | --- |
| TBD3 | NAI is an SRv6 segment | This Document |

## 4. Acknowledgements

The authors would like to thank Jeff Tentsura for suggestions regarding SRv6 MSD Types.

## 5. References

### 5.1. 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>.

[RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <https://www.rfc-editor.org/info/rfc5511>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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>.

[RFC8491] Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling Maximum SID Depth (MSD) Using IS-IS", RFC 8491, DOI 10.17487/RFC8491, November 2018, <https://www.rfc-editor.org/info/rfc8491>.

[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-14 (work in progress), October 2018.

## 5.2. Informative References

[RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <https://www.rfc-editor.org/info/rfc4657>.

[RFC7855] Previdi, S., Ed., Filsfils, C., Ed., Decraene, B., Litkowski, S., Horneffer, M., and R. Shakir, "Source Packet Routing in Networking (SPRING) Problem Statement and Requirements", RFC 7855, DOI 10.17487/RFC7855, May 2016, <https://www.rfc-editor.org/info/rfc7855>.

[RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <https://www.rfc-editor.org/info/rfc8051>.

[RFC8354] Brzozowski, J., Leddy, J., Filsfils, C., Maglione, R., Ed., and M. Townsley, "Use Cases for IPv6 Source Packet Routing in Networking (SPRING)", RFC 8354, DOI 10.17487/RFC8354, March 2018, <https://www.rfc-editor.org/info/rfc8354>.

[I-D.ietf-6man-segment-routing-header]
Filsfils, C., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-16 (work in progress), February 2019.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-22 (work in progress), December 2018.

[I-D.ietf-ospf-ospfv3-segment-routing-extensions]
Psenak, P. and S. Previdi, "OSPFv3 Extensions for Segment Routing", draft-ietf-ospf-ospfv3-segment-routing-extensions-23 (work in progress), January 2019.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d., bogdanov@google.com, b., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-02 (work in progress), October 2018.

[I-D.filsfils-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-filsfils-spring-srv6-network-programming-06 (work in progress), October 2018.

[I-D.li-ospf-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", draft-li-ospf-ospfv3-srv6-extensions-02 (work in progress), September 2018.

[I-D.bashandy-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Routing over IPv6 Dataplane", draft-bashandy-isis-srv6-extensions-04 (work in progress), October 2018.

[I-D.dawra-idr-bgpls-srv6-ext]
Dawra, G., Filsfils, C., Talaulikar, K., Chen, M., daniel.bernier@bell.ca, d., Uttaro, J., Decraene, B., and H. Elmalky, "BGP Link State extensions for IPv6 Segment Routing(SRv6)", draft-dawra-idr-bgpls-srv6-ext-04 (work in progress), September 2018.

## Appendix A. Contributor

The following persons contributed to this document:

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Huang Wumin
Huawei Technologies
Huawei Building, No. 156 Beiqing Rd.
Beijing 100095
China

Email: huangwumin@huawei.com

## Authors' Addresses

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahendrasingh@huawei.com

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

EMail: chengli13@huawei.com

Siva Sivabalan
Cisco Systems

EMail: msiva@cisco.com

Prejeeth Kaladharan
RtBrick Inc
Bangalore, Karnataka
India

EMail: prejeeth@rtbrick.com

PCE Working Group S. Sivabalan
Internet-Draft C. Filsfils
Intended status: Standards Track Cisco Systems, Inc.
Expires: April 22, 2019 J. Tantsura
Apstra, Inc.
J. Hardwick
Metaswitch Networks
S. Previdi
D. Dhody
Huawei Technologies
October 19, 2018

# Carrying Binding Label/Segment-ID in PCE-based Networks.
draft-sivabalan-pce-binding-label-sid-05

## Abstract

In order to provide greater scalability, network opacity, and service independence, SR utilizes a Binding Segment Identifier (BSID). It is possible to associate a BSID to RSVP-TE signaled Traffic Engineering Label Switching Path or binding Segment-ID (SID) to Segment Routed (SR) Traffic Engineering path. Such a binding label/SID can be used by an upstream node for steering traffic into the appropriate TE path to enforce SR policies. This document proposes an approach for reporting binding label/SID to Path Computation Element (PCE) for supporting PCE-based Traffic Engineering policies.

## Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## Status of This Memo

Table of Contents



## 1. Introduction

A PCE can compute Traffic Engineering paths (TE paths) through a network that are subject to various constraints. Currently, TE paths are either set up using the RSVP-TE signaling protocol or Segment Routing (SR). We refer to such paths as RSVP-TE paths and SR-TE paths respectively in this document.

As per [RFC8402] SR allows a headend node to steer a packet flow along any path. The headend node is said to steer a flow into an Segment Routing Policy (SR Policy). Further, as per [I-D.ietf-spring-segment-routing-policy], an SR Policy is a framework that enables instantiation of an ordered list of segments on a node for implementing a source routing policy with a specific intent for traffic steering from that node.

As described in [RFC8402], Binding Segment Identifier (BSID) is bound to an Segment Routed (SR) Policy, instantiation of which may involve a list of SIDs. Any packets received with an active segment equal to BSID are steered onto the bound SR Policy. A BSID may be either a local (SRLB) or a global (SRGB) SID. As per [I-D.ietf-spring-segment-routing-policy] a BSID can also be associated with any type of interfaces or tunnel to enable the use of a non-SR interface or tunnels as segments in a SID-list.

[RFC5440] describes the Path Computation Element Protocol (PCEP) for communication between a Path Computation Client (PCC) and a PCE or between a pair of PCEs. [RFC8231] specifies extension to PCEP that allows a PCC to delegate its LSPs to a stateful PCE. A stateful PCE can then update the state of LSPs delegated to it. [RFC8281] specifies a mechanism allowing a PCE to dynamically instantiate an LSP on a PCC by sending the path and characteristics. The PCEP extension to setup and maintain SR-TE paths is specified in [I-D.ietf-pce-segment-routing].

[I-D.ietf-pce-segment-routing] provides a mechanism for a network controller (acting as a PCE) to instantiate candidate paths for an SR Policy onto a head-end node (acting as a PCC) using PCEP. For more information on the SR Policy Architecture, see [I-D.ietf-spring-segment-routing-policy].

Binding label/SID has local significance to the ingress node of the corresponding TE path. When a stateful PCE is deployed for setting up TE paths, it may be desirable to report the binding label or SID to the stateful PCE for the purpose of enforcing end-to-end TE/SR policy. A sample Data Center (DC) use-case is illustrated in the following diagram. In the MPLS DC network, an SR LSP (without traffic engineering) is established using a prefix SID advertised by BGP (see [I-D.ietf-idr-bgp-prefix-sid]). In IP/MPLS WAN, an SR-TE LSP is setup using the PCE. The list of SIDs of the SR-TE LSP is {A, B, C, D}. The gateway node 1 (which is the PCC) allocates a binding SID X and reports it to the PCE. In order for the access node to steer the traffic over the SR-TE LSP, the PCE passes the SID stack {Y, X} where Y is the prefix SID of the gateway node-1 to the access node. In the absence of the binding SID X, the PCE should pass the SID stack {Y, A, B, C, D} to the access node. This example also illustrates the additional benefit of using the binding SID to reduce the number of SIDs imposed on the access nodes with a limited forwarding capacity.

```
           SID stack
           {Y, X}              +-----+
    _ _ _ _ _ _ _ _ _ _ _ _ _ _| PCE |
   |                           +-----+
   |                              ^
   |                              | Binding
   |           .-----.            | SID (X)     .-----.
   |          (       )           |            (       )
   V       .--(         )--.      |        .--(         )--.
+------+  (                 )  +-------+  (                 )  +-------+
|Access|_(  MPLS DC Network  )_|Gateway|_(    IP/MPLS WAN    )_|Gateway|
| Node | (  ==============>  ) |Node-1 | ( ================> ) |Node-2 |
+------+  (     SR path     )  +-------+  (    SR-TE path    )  +-------+
           '--(         )--'    Prefix     '--(         )--'
               (       )        SID of         (       )
                '-----'         Node-1          '-----'
                                is Y            SIDs for SR-TE LSP:
                                                {A, B, C, D}
```

Figure 1: A sample Use-case of Binding SID

A PCC could report the binding label/SID allocated by it to the stateful PCE via Path Computation State Report (PCRpt) message. It is also possible for a stateful PCE to request a PCC to allocate a specific binding label/SID by sending an Path Computation Update Request (PCUpd) message. If the PCC can successfully allocate the specified binding value, it reports the binding value to the PCE. Otherwise, the PCC sends an error message to the PCE indicating the cause of the failure. A local policy or configuration at the PCC SHOULD dictate if the binding label/SID needs to be assigned.

In this document, we introduce a new OPTIONAL TLV that a PCC can use in order to report the binding label/SID associated with a TE LSP, or a PCE to request a PCC to allocate a specific binding label/SID value. This TLV is intended for TE LSPs established using RSVP-TE, SR, or any other future method. Also, in the case of SR-TE LSPs, the TLV can carry a binding MPLS label (for SR-TE path with MPLS data-plane) or a binding IPv6 SID (e.g., IPv6 address for SR-TE paths with IPv6 data-plane). However, use of this TLV for carrying non-MPLS binding SID will be described in separate document(s). Binding value means either MPLS label or SID throughout this document.

## 2. Terminology

The following terminologies are used in this document:

BSID:  Binding Segment Identifier.

LER:  Label Edge Router.

LSP:  Label Switched Path.

LSR:  Label Switching Router.

PCC:  Path Computation Client.

PCE:  Path Computation Element

PCEP:  Path Computation Element Protocol.

RSVP-TE:  Resource ReserVation Protocol-Traffic Engineering.

SID:  Segment Identifier.

SR:  Segment Routing.

SRGB:  Segment Routing Global Block.

SRLB:  Segment Routing Local Block.

TLV:  Type, Length, and Value.

## 3. Path Binding TLV

The new optional TLV is called "TE-PATH-BINDING TLV" whose format is shown in the diagram below is defined to carry binding label or SID for a TE path. This TLV is associated with the LSP object specified in ([RFC8231]). The type of this TLV is to be allocated by IANA.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Binding Type (BT)        |        Binding Value          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~            Binding Value (continued) (variable length)        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: TE-PATH-BINDING TLV

TE-PATH-BINDING TLV is a generic TLV such that it is able to carry MPLS label binding as well as other types of future bindings (e.g., SRv6 path). It is formatted according to the rules specified in [RFC5440]. The two byte Binding Type (BT) field identifies the type of binding included in the TLV. This document specifies the following BT values:

- o BT = 0: The binding value is an MPLS label carried in the format specified in [RFC5462] where only the label value is valid, and other fields (TC, S, and TTL) fields MUST be considered invalid. The Length MUST be set to 6.

- o BT = 1: Similar to the case where BT is 0 except that all the fields on the MPLS label entry are set on transmission. However, the receiver MAY choose to override TC, S, and TTL values according its local policy.

Binding Value: A variable length field, padded with trailing zeros to a 4-byte boundary. For the BT as 0, the 20 bits represents the MPLS label. For the BT as 1, the 32-bits represents the label stack entry as per [RFC5462].

## 4. Operation

The binding value is allocated by the PCC and reported to a PCE via PCRpt message. If a PCE does not recognize the TE-PATH-BINDING TLV, it MUST ignore the TLV in accordance with ([RFC5440]). If a PCE recognizes the TLV but does not support the TLV, it MUST send PCErr with Error-Type = 2 (Capability not supported).

If a TE-PATH-BINDING TLV is absent in PCRpt message, PCE MUST assume that the corresponding LSP does not have any binding. If there are more than one TE-PATH-BINDING TLVs, only the first TLV MUST be processed and the rest MUST be silently ignored. If a PCE recognizes an invalid binding value (e.g., label value from the reserved label space when MPLS label binding is used), it MUST send the PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error Value = TBD ("Bad label value") as specified in [I-D.ietf-pce-segment-routing].

If a PCE requires a PCC to allocate a specific binding value, it may do so by sending a PCUpd or PCInitiate message containing a TE-PATH-BINDING TLV. If the value can be successfully allocated, the PCC reports the binding value to the PCE. If the PCC considers the binding value specified by the PCE invalid, it MUST send a PCErr message with Error-Type = TBD ("Binding label/SID failure") and Error Value = TBD ("Invalid SID"). If the binding value is valid, but the PCC is unable to allocate the binding value, it MUST send a PCErr message with Error-Type = TBD ("Binding label/SID failure") and Error Value = TBD ("Unable to allocate the specified label/SID").

If a PCC receives TE-PATH-BINDING TLV in any message other than PCUpd or PCInitiate, it MUST close the corresponding PCEP session with the reason "Reception of a malformed PCEP message" (according to [RFC5440]). Similarly, if a PCE receives a TE-PATH-BINDING TLV in any message other than a PCRpt or if the TE-PATH-BINDING TLV is associated with any object other than LSP object, the PCE MUST close the corresponding PCEP session with the reason "Reception of a malformed PCEP message" (according to [RFC5440]).

If a PCC wishes to withdraw or modify a previously reported binding value, it MUST send a PCRpt message without any TE-PATH-BINDING TLV or with the TE-PATH-BINDING TLV containing the new binding value respectively.

If a PCE wishes to modify a previously requested binding value, it MUST send a PCUpd message with TE-PATH-BINDING TLV containing the new binding value. Absence of TE-PATH-BINDING TLV in PCUpd message means that the PCE does not specify a binding value in which case the binding value allocation is governed by the PCC's local policy.

If a PCC receives a valid binding value from a PCE which is different than the current binding value, it MUST try to allocate the new value. If the new binding value is successfully allocated, the PCC MUST report the new value to the PCE. Otherwise, it MUST send a PCErr message with Error-Type = TBD ("Binding label/SID failure") and Error Value = TBD ("Unable to allocate the specified label/SID").

In some cases, a stateful PCE can request the PCC to allocate a binding value. It may do so by sending a PCUpd message containing an empty TE-PATH-BINDING TLV, i.e., no binding value is specified (making the length field of the TLV as 2). A PCE can also make the request PCC to allocate a binding at the time of initiation by sending a PCInitiate message with an empty TE-PATH-BINDING TLV.

## 5. Security Considerations

The security considerations described in [RFC5440], [RFC8231], [RFC8281] and [I-D.ietf-pce-segment-routing] are applicable to this specification. No additional security measure is required.

As described [I-D.ietf-pce-segment-routing], SR allows a network controller to instantiate and control paths in the network. Note that if the security mechanisms of [RFC5440] and [RFC8281] are not used, then the protocol described in this document could be attacked via manipulation of BSID.

# 6. IANA Considerations

## 6.1. PCEP TLV Type Indicators

This document defines a new PCEP TLV; IANA is requested to make the following allocations from the "PCEP TLV Type Indicators" sub-registry of the PCEP Numbers registry, as follows:

| Value | Name | Reference |
|---|---|---|
| TBD | TE-PATH-BINDING | This document |

### 6.1.1. TE-PATH-BINDING TLV

IANA is requested to create a sub-registry to manage the value of the Binding Type field in the TE-PATH-BINDING TLV.

| Value | Description | Reference |
|---|---|---|
| 0 | MPLS Label | This document |
| 1 | MPLS Label Stack Entry | This document |

## 6.2. PCEP Error Type and Value

This document defines a new Error-type and Error-Values for the PCErr message. IANA is requested to allocate new error-type and error-values within the "PCEP-ERROR Object Error Types and Values" subregistry of the PCEP Numbers registry, as follows:

| Error-Type | Meaning | |
|---|---|---|
| TBD | Binding label/SID failure: | |
| | Error-value = TBD: | Invalid SID |
| | Error-value = TBD: | Unable to allocate the specified label/SID |

# 7. Manageability Considerations

TBD

## 8. Acknowledgements

We like to thank Milos Fabian for his valuable comments.

## 9. References

### 9.1. 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>.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, DOI 10.17487/RFC5462, February 2009, <https://www.rfc-editor.org/info/rfc5462>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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-14 (work in progress), October 2018.

## 9.2. Informative References

[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>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d., bogdanov@google.com, b., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-01 (work in progress), June 2018.

[I-D.ietf-idr-bgp-prefix-sid]
Previdi, S., Filsfils, C., Lindem, A., Sreekantiah, A., and H. Gredler, "Segment Routing Prefix SID extensions for BGP", draft-ietf-idr-bgp-prefix-sid-27 (work in progress), June 2018.

[I-D.li-pce-sr-path-segment]
Li, C., Chen, M., Dhody, D., Cheng, W., Dong, J., Li, Z., and R. Gandhi, "Path Computation Element Communication Protocol (PCEP) Extension for Path Identification in Segment Routing (SR)", draft-li-pce-sr-path-segment-02 (work in progress), September 2018.

[I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-03 (work in progress), June 2018.

## Appendix A. PCE based Central Controller

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

As per [RFC8283], PCE as a central controller can allocate and provision the node/prefix/adjacency label (SID) via PCEP. It can also be used to allocate the binding SID as described in this section.

The PCECC Capability as per [I-D.zhao-pce-pcep-extension-pce-controller-sr] should also be advertised on the PCEP session, along with the SR sub-TLVs before using this procedure.

A P flag in LSP object is introduced in [I-D.li-pce-sr-path-segment] to indicate the allocation needs to be made by the PCE. The same flag is also set for the binding SID allocation request. A PCC would set this bit to 1 to request for allocation of the binding label/SID by the PCE in the PCReq or PCRpt message. A PCE would also set this bit to 1 to indicate that the binding label/SID is allocated by PCE and encoded in the PCRep, PCUpd or PCInitiate message (the TE-PATH-BINDING TLV is present in LSP object). Further, a PCE would set this bit to 0 to indicate that the path identifier is allocated by the PCC as described above.

The ingress PCC could request the binding label/SID to be allocated by the PCE via PCRpt message as per [RFC8231]. The delegate flag (D-flag) MUST also be set for this LSP. The TE-PATH-BINDING TLV MAY be included with no Binding Value. The PCECC would allocated the binding label/SID and further respond to Ingress PCC with PCUpd message as per [RFC8231] and MUST include the TE-PATH-BINDING TLV in a LSP object. The P flag in the LSP object would be set to 1 to indicate that the allocation is made by the PCE.

The PCE could allocate the binding label/SID on its own accord for a PCE- Initiated (or delegated LSP). The allocated binding label/SID needs to be informed to the PCC. The PCE would use the PCInitiate message [RFC8281] or PCUpd message [RFC8231] towards the PCC and MUST include the TE-PATH-BINDING TLV in the LSP object. The P flag in the LSP object would be set to 1 to indicate that the allocation is made by the PCE.

Authors' Addresses

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario  K2K 3E8
Canada

EMail: msiva@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Pegasus Parc
De kleetlaan 6a, DIEGEM  BRABANT 1831
BELGIUM

EMail: cfilsfil@cisco.com

Jeff Tantsura
Apstra, Inc.

EMail: jefftant.ietf@gmail.com

Jonathan Hardwick
Metaswitch Networks
100 Church Street
Enfield, Middlesex
UK

EMail: Jonathan.Hardwick@metaswitch.com

Stefano Previdi
Huawei Technologies

EMail: stefano@previdi.net

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 9, 2020

S. Sivabalan
C. Filsfils
Cisco Systems, Inc.
J. Tantsura
Apstra, Inc.
J. Hardwick
Metaswitch Networks
S. Previdi
C. Li
Huawei Technologies

July 8, 2019

# Carrying Binding Label/Segment-ID in PCE-based Networks.
draft-sivabalan-pce-binding-label-sid-07

## Abstract

In order to provide greater scalability, network opacity, and service independence, SR utilizes a Binding Segment Identifier (BSID). It is possible to associate a BSID to RSVP-TE signaled Traffic Engineering Label Switching Path or binding Segment-ID (SID) to Segment Routed (SR) Traffic Engineering path. Such a binding label/SID can be used by an upstream node for steering traffic into the appropriate TE path to enforce SR policies. This document proposes an approach for reporting binding label/SID to Path Computation Element (PCE) for supporting PCE-based Traffic Engineering policies.

## Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## Status of This Memo

Copyright Notice



Table of Contents

## 1. Introduction

A PCE can compute Traffic Engineering paths (TE paths) through a network that are subject to various constraints. Currently, TE paths are either set up using the RSVP-TE signaling protocol or Segment Routing (SR). We refer to such paths as RSVP-TE paths and SR-TE paths respectively in this document.

As per [RFC8402] SR allows a headend node to steer a packet flow along any path. The headend node is said to steer a flow into an Segment Routing Policy (SR Policy). Further, as per [I-D.ietf-spring-segment-routing-policy], an SR Policy is a framework that enables instantiation of an ordered list of segments on a node for implementing a source routing policy with a specific intent for traffic steering from that node.

As described in [RFC8402], Binding Segment Identifier (BSID) is bound to an Segment Routed (SR) Policy, instantiation of which may involve a list of SIDs. Any packets received with an active segment equal to BSID are steered onto the bound SR Policy. A BSID may be either a local (SRLB) or a global (SRGB) SID. As per Section 6.4 of [I-D.ietf-spring-segment-routing-policy] a BSID can also be associated with any type of interfaces or tunnel to enable the use of a non-SR interface or tunnels as segments in a SID-list.

[RFC5440] describes the Path Computation Element Protocol (PCEP) for communication between a Path Computation Client (PCC) and a PCE or between a pair of PCEs as per [RFC4655]. [RFC8231] specifies extension to PCEP that allows a PCC to delegate its LSPs to a stateful PCE. A stateful PCE can then update the state of LSPs delegated to it. [RFC8281] specifies a mechanism allowing a PCE to dynamically instantiate an LSP on a PCC by sending the path and characteristics. The PCEP extension to setup and maintain SR-TE paths is specified in [I-D.ietf-pce-segment-routing].

[I-D.ietf-pce-segment-routing] provides a mechanism for a network controller (acting as a PCE) to instantiate candidate paths for an SR Policy onto a head-end node (acting as a PCC) using PCEP. For more information on the SR Policy Architecture, see [I-D.ietf-spring-segment-routing-policy].

Binding label/SID has local significance to the ingress node of the corresponding TE path. When a stateful PCE is deployed for setting up TE paths, it may be desirable to report the binding label or SID to the stateful PCE for the purpose of enforcing end-to-end TE/SR policy. A sample Data Center (DC) use-case is illustrated in the following diagram. In the MPLS DC network, an SR LSP (without traffic engineering) is established using a prefix SID advertised by

BGP (see [I-D.ietf-idr-bgp-prefix-sid]). In IP/MPLS WAN, an SR-TE LSP is setup using the PCE. The list of SIDs of the SR-TE LSP is {A, B, C, D}. The gateway node 1 (which is the PCC) allocates a binding SID X and reports it to the PCE. In order for the access node to steer the traffic over the SR-TE LSP, the PCE passes the SID stack {Y, X} where Y is the prefix SID of the gateway node-1 to the access node. In the absence of the binding SID X, the PCE should pass the SID stack {Y, A, B, C, D} to the access node. This example also illustrates the additional benefit of using the binding SID to reduce the number of SIDs imposed on the access nodes with a limited forwarding capacity.

```
           SID stack
           {Y, X}              +-----+
    _ _ _ _ _ _ _ _ _ _ _ _ _ _| PCE |
   |                           +-----+
   |                              ^
   |                              | Binding
   |           .-----.            | SID (X)     .-----.
   |          (       )           |            (       )
   V       .--(         )--.      |        .--(         )--.
+------+  (                 )  +-------+  (                 )  +-------+
|Access|_(  MPLS DC Network  )_|Gateway|_(    IP/MPLS WAN    )_|Gateway|
| Node | (  ==============>  ) |Node-1 | ( ================> ) |Node-2 |
+------+  (     SR path     )  +-------+  (    SR-TE path   )  +-------+
           '--(         )--'    Prefix     '--(         )--'
               (       )        SID of         (       )
                '-----'         Node-1          '-----'
                                is Y            SIDs for SR-TE LSP:
                                                {A, B, C, D}
```

Figure 1: A sample Use-case of Binding SID

A PCC could report the binding label/SID allocated by it to the stateful PCE via Path Computation State Report (PCRpt) message. It is also possible for a stateful PCE to request a PCC to allocate a specific binding label/SID by sending an Path Computation Update Request (PCUpd) message. If the PCC can successfully allocate the specified binding value, it reports the binding value to the PCE. Otherwise, the PCC sends an error message to the PCE indicating the cause of the failure. A local policy or configuration at the PCC SHOULD dictate if the binding label/SID needs to be assigned.

In this document, we introduce a new OPTIONAL TLV that a PCC can use in order to report the binding label/SID associated with a TE LSP, or a PCE to request a PCC to allocate a specific binding label/SID value. This TLV is intended for TE LSPs established using RSVP-TE, SR, or any other future method. Also, in the case of SR-TE LSPs, the TLV can carry a binding MPLS label (for SR-TE path with MPLS data-plane) or a binding IPv6 SID (e.g., IPv6 address for SR-TE paths with IPv6 data-plane). Binding value means either MPLS label or SID throughout this document.

Additionally, to support the PCE based central controller [RFC8283] operation where the PCE would take responsibility for managing some part of the MPLS label space for each of the routers that it controls, the PCE could directly make the binding label/SID allocation and inform the PCC. See [I-D.ietf-pce-pcep-extension-for-pce-controller] for details.

## 2. Terminology

The following terminologies are used in this document:

BSID: Binding Segment Identifier.

LER: Label Edge Router.

LSP: Label Switched Path.

LSR: Label Switching Router.

PCC: Path Computation Client.

PCE: Path Computation Element

PCEP: Path Computation Element Protocol.

RSVP-TE: Resource ReserVation Protocol-Traffic Engineering.

SID: Segment Identifier.

SR: Segment Routing.

SRGB: Segment Routing Global Block.

SRLB: Segment Routing Local Block.

TLV: Type, Length, and Value.

## 3. Path Binding TLV

The new optional TLV is called "TE-PATH-BINDING TLV" whose format is shown in the diagram below is defined to carry binding label or SID for a TE path. This TLV is associated with the LSP object specified in ([RFC8231]). The type of this TLV is to be allocated by IANA.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Type              |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      BT       |                 Reserved                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~            Binding Value (variable length)                    ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: TE-PATH-BINDING TLV

TE-PATH-BINDING TLV is a generic TLV such that it is able to carry MPLS label binding as well as SRV6 Binding SID. It is formatted according to the rules specified in [RFC5440].

Binding Type (BT): A one byte field identifies the type of binding included in the TLV. This document specifies the following BT values:

- o BT = 0: The binding value is an MPLS label carried in the format specified in [RFC5462] where only the label value is valid, and other fields (TC, S, and TTL) fields MUST be considered invalid. The Length MUST be set to 6.

- o BT = 1: Similar to the case where BT is 0 except that all the fields on the MPLS label entry are set on transmission. However, the receiver MAY choose to override TC, S, and TTL values according its local policy.

- o BT = 2: The binding value is a SRv6 SID with a format of an 16 byte IPv6 address, representing the binding SID for SRv6.

Reserved: MUST be set to 0 while sending and ignored on receipt.

Binding Value: A variable length field, padded with trailing zeros to a 4-byte boundary. For the BT as 0, the 20 bits represents the MPLS label. For the BT as 1, the 32-bits represents the label stack entry as per [RFC5462]. For the BT as 2, the 128-bits represent the SRv6 SID.

## 4. Operation

The binding value is allocated by the PCC and reported to a PCE via PCRpt message. If a PCE does not recognize the TE-PATH-BINDING TLV, it MUST ignore the TLV in accordance with ([RFC5440]). If a PCE recognizes the TLV but does not support the TLV, it MUST send PCErr with Error-Type = 2 (Capability not supported).

If a TE-PATH-BINDING TLV is absent in PCRpt message, PCE MUST assume that the corresponding LSP does not have any binding. If there are more than one TE-PATH-BINDING TLVs, only the first TLV MUST be processed and the rest MUST be silently ignored. If a PCE recognizes an invalid binding value (e.g., label value from the reserved label space when MPLS label binding is used), it MUST send the PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error Value = TBD ("Bad label value") as specified in [I-D.ietf-pce-segment-routing].

If a PCE requires a PCC to allocate a specific binding value, it may do so by sending a PCUpd or PCInitiate message containing a TE-PATH-BINDING TLV. If the value can be successfully allocated, the PCC reports the binding value to the PCE. If the PCC considers the binding value specified by the PCE invalid, it MUST send a PCErr message with Error-Type = TBD ("Binding label/SID failure") and Error Value = TBD ("Invalid SID"). If the binding value is valid, but the PCC is unable to allocate the binding value, it MUST send a PCErr message with Error-Type = TBD ("Binding label/SID failure") and Error Value = TBD ("Unable to allocate the specified label/SID").

If a PCC receives TE-PATH-BINDING TLV in any message other than PCUpd or PCInitiate, it MUST close the corresponding PCEP session with the reason "Reception of a malformed PCEP message" (according to [RFC5440]). Similarly, if a PCE receives a TE-PATH-BINDING TLV in any message other than a PCRpt or if the TE-PATH-BINDING TLV is associated with any object other than LSP object, the PCE MUST close the corresponding PCEP session with the reason "Reception of a malformed PCEP message" (according to [RFC5440]).

If a PCC wishes to withdraw or modify a previously reported binding value, it MUST send a PCRpt message without any TE-PATH-BINDING TLV or with the TE-PATH-BINDING TLV containing the new binding value respectively.

If a PCE wishes to modify a previously requested binding value, it MUST send a PCUpd message with TE-PATH-BINDING TLV containing the new binding value. Absence of TE-PATH-BINDING TLV in PCUpd message means that the PCE does not specify a binding value in which case the binding value allocation is governed by the PCC's local policy.

If a PCC receives a valid binding value from a PCE which is different than the current binding value, it MUST try to allocate the new value. If the new binding value is successfully allocated, the PCC MUST report the new value to the PCE. Otherwise, it MUST send a PCErr message with Error-Type = TBD ("Binding label/SID failure") and Error Value = TBD ("Unable to allocate the specified label/SID").

In some cases, a stateful PCE can request the PCC to allocate a binding value. It may do so by sending a PCUpd message containing an empty TE-PATH-BINDING TLV, i.e., no binding value is specified (making the length field of the TLV as 2). A PCE can also make the request PCC to allocate a binding at the time of initiation by sending a PCInitiate message with an empty TE-PATH-BINDING TLV.

## 5. Binding SID in SR-ERO

In PCEP messages, LSP route information is carried in the Explicit Route Object (ERO), which consists of a sequence of subobjects. [I-D.ietf-pce-segment-routing] defines a new ERO subobject "SR-ERO subobject" capable of carrying a SID as well as the identity of the node/adjacency (NAI) represented by the SID. The NAI Type (NT) field indicates the type and format of the NAI contained in the SR-ERO. In case of binding SID, the NAI MUST NOT be included and NT MUST be set to zero. So as per Section 5.2.1 of [I-D.ietf-pce-segment-routing], for NT=0, the F bit MUST be 1, the S bit needs to be zero and the Length MUST be 8. Further the M bit MUST be set. If these conditions are not met, the entire ERO MUST be considered invalid and a PCErr message is sent with Error-Type = 10 ("Reception of an invalid object") and Error-Value = 11 ("Malformed object").

## 6. Binding SID in SRv6-ERO/

[I-D.ietf-pce-segment-routing] defines a new ERO subobject "SRv6-ERO subobject" for SRv6 SID. The NAI MUST NOT be included and NT MUST be set to zero. So as per Section 5.2.1 of [I-D.ietf-pce-segment-routing], for NT=0, the F bit MUST be 1, the S bit needs to be zero and the Length MUST be 24. If these conditions are not met, the entire ERO is considered invalid and a PCErr message is sent with Error-Type = 10 ("Reception of an invalid object") and Error-Value = 11 ("Malformed object") (as per [I-D.ietf-pce-segment-routing]).

## 7. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to RFC 7942.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

### 7.1. Huawei

- o Organization: Huawei
- o Implementation: Huawei's Router and Controller
- o Description: An experimental code-point is used and plan to request early code-point allocation from IANA after WG adoption.
- o Maturity Level: Production
- o Coverage: Full
- o Contact: mahendrasingh@huawei.com

## 8. Security Considerations

The security considerations described in [RFC5440], [RFC8231], [RFC8281] and [I-D.ietf-pce-segment-routing] are applicable to this specification. No additional security measure is required.

As described [I-D.ietf-pce-segment-routing], SR allows a network controller to instantiate and control paths in the network. A rouge PCE can manipulate binding SID allocations to move traffic around for some other LSPs that uses BSID in its SR-ERO.

Thus, as per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253], as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

## 9. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] apply to PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

### 9.1. Control of Function and Policy

A PCC implementation SHOULD allow the operator to configure the policy based on which PCC needs to allocates the binding label/SID.

### 9.2. Information and Data Models

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to include policy configuration for binding label/SID allocation.

### 9.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

### 9.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing].

### 9.5. Requirements On Other Protocols

Mechanisms defined in this document do not imply any new requirements on other protocols.

### 9.6. Impact On Network Operations

Mechanisms defined in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] also apply to PCEP extensions defined in this document. Further, the mechanism described in this document can help the operator to request control of the LSPs at a particular PCE.

## 10. IANA Considerations

### 10.1. PCEP TLV Type Indicators

This document defines a new PCEP TLV; IANA is requested to make the following allocations from the "PCEP TLV Type Indicators" sub-registry of the PCEP Numbers registry, as follows:

| Value | Name | Reference |
|---|---|---|
| TBD | TE-PATH-BINDING | This document |

#### 10.1.1. TE-PATH-BINDING TLV

IANA is requested to create a sub-registry to manage the value of the Binding Type field in the TE-PATH-BINDING TLV.

| Value | Description | Reference |
|---|---|---|
| 0 | MPLS Label | This document |
| 1 | MPLS Label Stack Entry | This document |
| 2 | SRv6 SID | This document |

### 10.2. PCEP Error Type and Value

This document defines a new Error-type and Error-Values for the PCErr message. IANA is requested to allocate new error-type and error-values within the "PCEP-ERROR Object Error Types and Values" subregistry of the PCEP Numbers registry, as follows:

| Error-Type | Meaning | |
|---|---|---|
| TBD | Binding label/SID failure: | |
| | Error-value = TBD: | Invalid SID |
| | Error-value = TBD: | Unable to allocate the specified label/SID |

## 11. Acknowledgements

We like to thank Milos Fabian for his valuable comments.

# 12. References

## 12.1. 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>.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, DOI 10.17487/RFC5462, February 2009, <https://www.rfc-editor.org/info/rfc5462>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <https://www.rfc-editor.org/info/rfc7525>.

[RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <https://www.rfc-editor.org/info/rfc7942>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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>.

[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>.

[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>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[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-16 (work in progress), March 2019.

## 12.2. Informative References

[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>.

[RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <https://www.rfc-editor.org/info/rfc8283>.

[I-D.ietf-spring-segment-routing-policy] Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d., bogdanov@google.com, b., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-03 (work in progress), May 2019.

[I-D.ietf-idr-bgp-prefix-sid] Previdi, S., Filsfils, C., Lindem, A., Sreekantiah, A., and H. Gredler, "Segment Routing Prefix SID extensions for BGP", draft-ietf-idr-bgp-prefix-sid-27 (work in progress), June 2018.

[I-D.ietf-pce-pcep-extension-for-pce-controller] Zhao, Q., Li, Z., Negi, M., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-ietf-pce-pcep-extension-for-pce-controller-01 (work in progress), February 2019.

[I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-12 (work in progress), July 2019.

## Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahendrasingh@huawei.com

## Authors' Addresses

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario K2K 3E8
Canada

EMail: msiva@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Pegasus Parc
De kleetlaan 6a, DIEGEM BRABANT 1831
BELGIUM

EMail: cfilsfil@cisco.com

Jeff Tantsura
Apstra, Inc.

EMail: jefftant.ietf@gmail.com

Jonathan Hardwick
Metaswitch Networks
100 Church Street
Enfield, Middlesex
UK

EMail: Jonathan.Hardwick@metaswitch.com

Stefano Previdi
Huawei Technologies

EMail: stefano@previdi.net

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

EMail: chengli13@huawei.com

PCE WG
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2019
Quan. Xiong
Fangwei. Hu
Ran. Chen
ZTE Corporation
October 22, 2018

# PCE Multi-layer LSP Association
draft-xiong-pce-multilayer-lsp-association-02

## Abstract

The Path Computation Element Communication Protocol (PCEP) provides mechanisms for Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients (PCCs) requests. [I-D.ietf-pce-association-group] proposed an association mechanism for a set of LSPs.

This document proposes a set of extensions to PCEP to associate a grouping of multi-layer LSPs. The extensions define a mechanism to create associations between upper-layer LSP and related lower-layer LSPs.

Table of Contents



## 1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) which is used between a Path Computation Element (PCE) and a Path Computation Client (PCC) (or other PCE) to enable computation of Multi-protocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP). [I-D.ietf-pce-association-group] proposed an association mechanism to create a grouping of LSPs in the context of a PCE.

This document proposes a set of extensions to PCEP to associate a grouping of multi-layer LSPs. The extensions define a mechanism to create associations between upper-layer LSP and related lower-layer LSPs.

## 2. 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 [RFC2119].

### 2.1. Terminology

The terminology is defined as [RFC5440] ,[I-D.ietf-pce-stateful-pce-app] and [I-D.ietf-pce-association-group].

## 3. Overview

### 3.1. Motivation

In GMPLS/MPLS networks, service provider network is divided into several service layers according to the requirements and customer network is the upper layer with the lower layers as the Forwarding Adjacency LSP (FA-LSP) as shown in Figure 1. The service connection is established with the set up of multi-layer LSPs.

```
        Initiate & Update LSP and related lower LSPs
                                   |
                                   |
                                   V
                               +---+--+
            +--------------+ PCE  +---------------+
            |                  +---+--+               |
            |                      |                  |
            |                      |                  |
            |                      |                  |
        +---+--+           +---+--+           +---+--+
        | PCC1 +-----------+ PCC2 +-----------+ PCC3 +
        +------+   LSP1->  +------+  LSP2->   +------+
            |                                         |
            |<------          LSP3           ----->|
```

Figure 1 Usecase for multi-layer LSPs

As discussed in [I-D.ietf-pce-stateful-pce-app] , it consists of a set of one or more TE LSPs in the lower layer which provides TE links to the upper layer in Multi-Layer Networks (MLN). The requirement is to control of the multi-layer LSPs and related TE links. The establishment or teardown of a lower layer LSP needs to take into consideration the state of existing LSPs or new LSP request in the upper layer.

As discussed in [I-D.ietf-pce-stateful-pce] , the stateful PCE MAY determine to optimize the link and path based on the lower layer of the LSP and its upper TE Link, and in the case of the failure of the lower level LSP, it MAY update the upper network LSP path according to the existing resources and the status of the LSP.

The stateful PCE provides the ability to update the LSP, in the process of bandwidth adjustment, it MAY be necessary to adjust the bandwidth of related lower layer LSPs, which provide the TE link for the upper layer LSP. The association of multi-layer LSPs can reduce the repeated operations and optimize the information interaction between PCC and PCE.

In overlayer multi-domain scenario, the lower-layer LSPs in each domain may be initiated by respective domain’s PCE and stitched together to an association group with an end-to-end LSP as its upper-layer LSP.

In these cases, it is necessary to add multi-layer LSPs to an association group.

## 3.2. Operation Overview

[I-D.ietf-pce-association-group] introduces a generic mechanism to create a grouping of LSPs. This grouping can then be used to define associations between sets of LSPs or between a set of LSPs and a set of attributes.

In order to solve the problem of multi-layer LSP control in PCE network, this document proposes the association if the multi-layer LSPs. The upper LSP is associated with its related lower LSPs by adding them to a multi-layer association group.

One new optional Association Object type is defined carried in the Association object defined in [I-D.ietf-pce-association-group]. This document proposes a new association type called "Layer Association Type" and related TLV called "LAYER-ASSOCIATION TLV".

As defined in [I-D.ietf-pce-association-group], multi-layer LSPs associations could be created dynamically or configured by the operator when operator-configured association is needed.

The handling and policy of multi-layer LSPs Association is similar to the generic association and some processing rules as shown in session 4.2.

## 4. Extensions to the PCEP

### 4.1. Association Type and Group

[I-D.ietf-pce-association-group] introduces the ASSOCIATION object and this document proposes a new Association type for multi-layer LSPs association to associate multi-layer LSPs into one group for further operation. An association ID will be used to identify the group and a new Association Type is defined in this document, based on the generic Association object :

Association type = TBD1 ("Multi-Layer Association Type") for Multi-Layer Association Group (MLAG)

MLAG may carry optional TLVs including but not limited to :

MULTI-LAYER-ASSOCIATION-TLV: Used to identify the upper-layer LSP and lower-layer LSP in multi-layer information, described in Section 4.2.

As [I-D.ietf-pce-association-group] specified, the capability advertisement of the association types supported by a PCEP speaker is performed by defining a ASSOC-Type-List TLV to be carried within an OPEN object. The association type which defined in this document should be added in the list and be advertised between the PCEP speakers before the multi-layer association.

This Association-Type is operator-configured and created by the operator manually on the PCEP peers. The LSP belonging to this associations is conveyed via PCEP messages to the PCEP peer. Operator-configured Association Range SHOULD NOT be set for this association-type, and MUST be ignored, so that the full range of association identifier can be utilized.

### 4.2. MULTI-LAYER-ASSOCIATION TLV

This document proposes LAYER-ASSOCIATION TLV for the association of multi-layer LSPs. The TLV is optional. The format of the new Association TLV is shown in Figure 4:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Type = TBD        |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Layer Association Flags                   |H|L|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: The LAYER-ASSOCIATION TLV format

The type of the TLV is [TBD] which indicates the LAYER ASSOCIATION TLV. The fields in the format are:

Length:16bits,the length of the TLV.

Layer Association Flags-H:1bit, indicates LSP of the upper layer when it is set.

Layer Association Flags-L:1bit, indicates LSP of the lower layer when it is set.

## 5. PCEP Procedure

Once a group of multilayer LSPs is created, the upper layer LSP is associated with its related lower layer LSPs. Association objects can be carried in PCReq, PCRpt, PCUpd, or PCInit messages.

### 5.1. Multi-Layer LSPs Associations Creation

As defined in [I-D.ietf-pce-association-group], association groups can be created by both PCC and PCE.

In stateless PCE, the association object with "Layer Association type" is carried in PCReq message from PCC to PCE, indicating that the LSP joins one existing multi-layer LSPs association group or create a new one. If the LSP is belong to upper layer then set the "H" bit in "LAYER-ASSOCIATION TLV", otherwise set the "L" bit when it is lower layer LSP.

In stateful PCE, PCE MAY create a new association group or associate a LSP to an existing association group carried in PCInit message after the LSP delegation from PCC to the PCE as discussed in [I-D.ietf-pce-pce-initiated-lsp]. In state synchronization process between PCC and PCE, PCC also need to report the existing multi-layer LSPs association groups to PCE. If the association group changes, PCC needs to report the relevant group changes to PCE through the PCRpt message.

### 5.2. Bandwidth Adjustment

The stateful PCE provides the ability to update the LSP, in the process of bandwidth adjustment, for example, enlarge the bandwidth of the upper layer LSP, it MUST be necessary to adjust the bandwidth of related lower layer LSPs, which provide the TE link for it.

Once the multi-layer LSPs associated in a group, the PCE MAY send the PCUpd message to the PCC with the association object to adjust the upper layer LSP. Once receiving the request, PCC will search the relevant lower layer LSPs and adjust their bandwidth before the adjustment of the upper layer LSP.

### 5.3. TE Links Optimization

The stateful PCE MAY determine to optimize the link and path based on the lower layer of the LSP and its upper TE Link, and in the case of the failure of the lower level LSP, it MAY update the upper network LSP path and re-optimize resource usage across multi-layers.

When removing the upper layer LSP, PCC or PCE MAY release each of lower layer LSPs which associated in a group and re-use the resources for other upper layer LSP according to the existing resources and the status of the LSP.

## 6. Security Considerations

TBD

## 7. IANA Considerations

### 7.1. Association Object Type

This document defines a new association type in Association object which originally defined in [I-D.ietf-pce-association-group]. IANA is requested to make allocations from the registry, as follows:

| Value | Name | Reference |
|---|---|---|
| TBD | Layer Association Type | [this document] |

Table 1

### 7.2. LAYER-ASSOCIATION TLV

This document defines the following TLV in Association object which originally defined in [I-D.ietf-pce-association-group]. IANA is requested to make allocations from the registry, as follows:

| Value | Name | Reference |
|---|---|---|
| TBD | LAYER-ASSOCIATION TLV | [this document] |

Table 2

## 8. Acknowledgements

TBD.

## 9. References

### 9.1. Informative References

[I-D.ietf-pce-stateful-pce-app]
Zhang, X. and I. Minei, "Applicability of a Stateful Path Computation Element (PCE)", draft-ietf-pce-stateful-pce-app-08 (work in progress), October 2016.

### 9.2. Normative References

[I-D.ietf-pce-association-group]
Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-06 (work in progress), June 2018.

[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.

[I-D.ietf-pce-stateful-pce]
Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-21 (work in progress), June 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>.

[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>.

## Authors' Addresses

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei  430223
China

Phone: +86 27 83531060
Email: xiong.quan@zte.com.cn

Fangwei Hu
ZTE Corporation
No.889 Bibo Rd
Shanghai  201203
China

Phone: +86 21 68896273
Email: hu.fangwei@zte.com.cn

Ran Chen
ZTE Corporation
No.50 Software Avenue,Yuhuatai District
Nanjing, Jiangsu Province  210012
China

Phone: +86 025 88014636
Email: chen.ran@zte.com.cn