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.

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 22, 2019.

Copyright Notice

   Copyright (c) 2018 IETF Trust and the persons identified as the
   document authors. All rights reserved.

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

Table of Contents

1. 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}
```
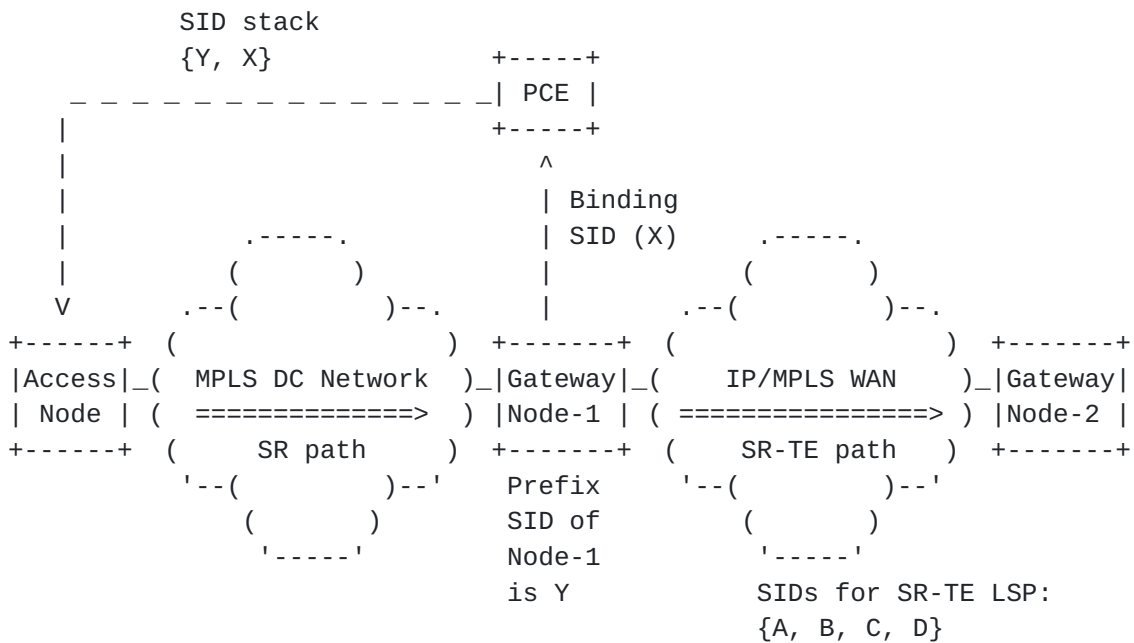
Figure 1: A sample Use-case of Binding SID

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)         ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
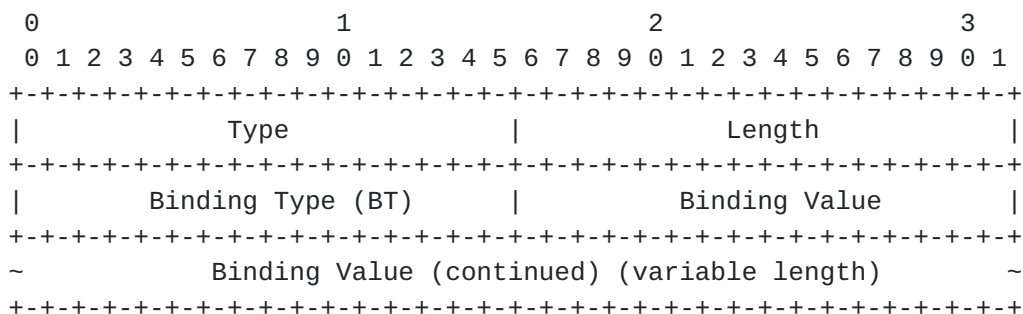
Figure 2: TE-PATH-BINDING TLV

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