pce                                                    H. Li
Internet-Draft                                      A. Wang
Intended status: Standards Track            China Telecom
Expires: September 8, 2022                       Z. Zhang
                                           Juniper Networks
                                                   H. Chen
                                                 Futurewei
                                                   R. Chen
                                           ZTE Corporation
                                            March 07, 2022

# Multicast Tree Setup via PCEP
## draft-li-pce-multicast-00

Abstract

   Multicast forwarding requires per-tree state on certain nodes.  Even
   with BIER, per-tree state is needed on ingress/egress BIER routers
   (though not needed on transit BIER routers).  This documents
   specifies PCEP protocol to collect tree information (e.g. root, leaf,
   constraints) to allow a PCE to calculate a tree, and the procedures
   to set up per-tree forwarding state on relevant nodes for various
   multicast trees and various replication technologies.

Status of This Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

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

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

   This Internet-Draft will expire on September 8, 2022.

Table of Contents

## 1. Introduction

Currently, multicast management information is mainly signaled by PIM
[RFC2362], mLDP [RFC6388] or BGP [RFC6514].  The trees/tunnels are
set up using the "receiver-initiated join" technique of PIM/mLDP, hop
by hop from downstream routers towards the root.  The BGP messages
are either sent hop by hop between downstream routers and their
upstream neighbors, or can be reflected by Route Reflectors (RRs).
The signaling is the interaction between routers and cannot be
managed in a unified manner.

As an alternative to each hop independently determining its upstream
router and signaling upstream towards the root (following PIM/mLDP
model), the entire tree can be calculated by a centralized
controller, and the signaling can be entirely done from the
controller.

[RFC4655] defines a stateful PCE to be one in which the PCE maintains
"strict synchronization between the PCE and not only the network
states (in term of topology and resource information), but also the
set of computed paths and reserved resources in use in the network."
[RFC8231] specifies a set of extensions to PCEP to support state
synchronization between PCCs and PCEs.

The controller can serve as a PCE to centrally manage multicast trees
and establish states on all tree nodes serving as PCC.  This document
defines PCEP objects and TLVs to accomplish the multicast source
registration/revocation, receiver discovery, multicast tree state
update etc. functions.

## 2.  Overview

### 2.1.  Different Multicast Trees

This section discusses various multicast trees like IP, mLDP, SR-P2MP - each with its own tree identification.

#### 2.1.1.  IP Multicast

As stated in [RFC1112], an IP multicast packet's destination address is a Multicast Group Address and it follows either a source specific tree rooted at the First Hop Router (FHR) or a shared tree rooted at a Rendezvous Point (RP) [RFC7761].  Each tree is identified by a (source address, group address) pair, where the source address MAY be a wildcard in case of the shared tree.  This identification, which is often referred to as (s, g) or (*, g), is used in both forwarding plane as well as in control plane that sets up the tree, e.g., using Protocol Independ Multicast (PIM) [RFC7761] [RFC5015].

Each node on the tree needs to have corresponding forwarding plane state used to forward corresponding traffic and control plane state that is used to set up a tree.  Depending on the number of (s, g) and/or (*, g) trees that a routing domain need to support, serious scaling/convergence problems MAY result.

#### 2.1.2.  mLDP/RSVP-TE P2MP Tunnels

Quite often, some different multicast trees are used for similar purposes and the tree structure are identical or very similar.  For example, in an IPTV application, many channels MAY need to be sent from the same headend router to the same set of edge routers close to receivers.

In that case, a set of IP multicast trees can be transported over a fewer set of MPLS P2MP tunnels - either mLDP [RFC6388] or RSVP-TE P2MP [RFC4875] tunnels - across an MPLS domain.  Inside the MPLS domain, there is no IP multicast tree state but only fewer state for the P2MP tunnels.  Outside the MPLS domain, those IP multicast trees still have their per-tree state on each tree node.

An mLDP tree is identified by an mLDP FEC in the control plane.  In the context of PCEP signaling, part of or an entire mLDP tunnel can be set up using PCEP from a PCE instead of using hop-by-hop mLDP signaling.  In this case the only relevance to mLDP is that mLDP FEC is used to identify the tunnel.

A RSVP-TE P2MP tree is identified by a RSVP Session object in the control plane.  While in theory it can also be set up via PCEP signaling from a PCE, it is outside the scope of this document.

In the forwarding plane, there is no difference between an mLDP tunnel and a RSVP-TE P2MP tunnel.  A tree is identified by a label - incoming labeled traffic is replicated to a bunch of downstream nodes with a label that identifies the tree to the downstream nodes.

### 2.1.3.  SR-P2MP Tunnels

Segment Routing (SR) [RFC8402] has two principles:

1.  No per-flow/tunnel state inside an SR domain

2.  Optional but perferred use of controllers

When it comes to multicast, the only technologies that sticks to principle #1 is BIER [RFC8279], which does not use per-tree state for forwarding.

SR-P2MP [I-D.ietf-pim-sr-p2mp-policy] [I-D.ietf-spring-sr-replication-segment] sticks to principle #2 only, in that PCEs are used to calculate a multicast tree subject to constraints and then direct signaling from PCEs are used to set up the tree instead of using mLDP/RSVP signaling, but per-tree state is still used.

In the control plane, an SR-P2MP is identified by a (root-id, tree-id) tuple.  In an SR-MPLS forwarding plane, an SR-P2MP tunnel is not different from mLDP/RSVP-TE P2MP tunnel at all, though the tree-identifying label is called a tree sid.

In an SRv6 forwarding plane, the tree sid is embedded as part of an IPv6 address.

### 2.2.  Different Replication Technologies

For any kind of multicast trees mentioned above, on a particular tree node A, an incoming packet needs to replicated to a set of downstream nodes B/C/D/E.  Note that the upstream and downstream nodes MAY be directly connected or they can be reached via a set of P2P/MP2P tunnels, P2MP tunnels, or via BIER.  In the latter case, intermediate nodes do not maintain the per-tree state but they do need to maintain state for the transporting tunnels or BIER.

Now on node A/B/C/D/E, per-tree replication state needs to be set up. In particular on A, depending on the method used to replicate to

B/C/D/E, (a combination of) the following replication branches are
needed:

o  A set of outgoing interfaces for native IP forwarding to directly
   connected nodes from the B/C/D/E set in case of IP multicast tree,
   and/or,

o  A set of P2P/MP2P/P2MP tunnels to reach indirectly connected nodes
   from the B/C/D/E set, and/or,

o  A BIER bitstring and relevant BIER information to reach some nodes
   from the B/C/D/E set

In the latter two cases, the replication branches also need relevant
information to identify to B/C/D/E which tree a packet is for (e.g.,
a label for an mLDP or SR-P2MP tree).

## 2.3.  Tree Information Discovery

When a PCE calculates a multicast tree, it needs to collect the tree
information like root, leaves and calculation constraints.  This MAY
be provided to the PCE via a southbound (wrt the PCE) interface from
an orchestrator or via northbound PCEP signaling from some PCC nodes
(e.g. the root and leaf nodes of the tree).  The PCE MAY also
participate overlay signaling protocols to collect the information
(e.g., [I-D.zzhang-mvpn-evpn-controller]).

This document only covers the northbound PCEP signaling.  The two
other methods mentioned above are out of scope of this document.

Whether it is an IP Multicast or mLDP/SR-P2MP Tree, the root, leaves
and calculation constraints are encoded the same way but associated
with different tree identifications in the PCEP signaling.
Specifically, the root and leaves are encoded as IP addresses.

The leaf information can be encoded as a full set of leaves or as
addition/removal delta.

## 2.3.1.  Root node Discovery

When a multicast source accesses to a First Hop Router (FHR), the FHR
originates a PCRpt message carrying the Multicast Source Registration
(MSR) object defined in Section 3.3, which provides FHR information.
FHR, as PCC, delegate the controller as PCE to calculate multicast
tree.  In the sent PCRpt message, the D bit of LSP Object is set to
1.

### 2.3.2.  Leaf node Discovery

   For IP multicast, Last Hop Router (LHR) MAY convert the multicast
   source and group address carried in the IGMP Join/Leave messages sent
   by local hosts and VPN information corresponding to local hosts into
   PCRpt messages and report these information to the controller.
   Similar to IGMP messages, PIM messages from other domains can also be
   used as receiver information and be converted into PCRpt messages to
   be reported to the controller.

   For mLDP and SR-P2MP multicast, The corresponding tree identifiers
   can also be carried in PCRpt messages and reported to the controller.

   When the first receiver accesses a Last Hop Router (LHR) to join a
   multicast tree, or when the last receiver leaves the LHR, the LHR
   originates a PCRpt message carrying the Multicast Receiver
   Information (MRI) object defined in Section 3.4, which provides LHR
   information.

   If the receivers locate in different VPN domains, the VPN information
   associated with the multicast source and multicast destination should
   also be reported by the LHR and FHR.  The VPN information reported by
   LHRs SHOULD be consistent.

### 2.4.  Multicast Tree

   The multicast trees are established with the FHRs of the multicast
   source as the root of the multicast trees.  According to different
   multicast technologies, the types of multicast trees include mLDP
   tree, SR-P2MP tree and BIER tree.  Different multicast trees
   correspond to different processing.  This section describes the state
   update of different multicast operations.

   No matter which forwarding technology is used, there is a case that
   local multicast receivers directly access to FHRs.  In this case,
   FHRs need to forward multicast data for local receivers in the way of
   native IP without other control information.

### 2.4.1.  mLDP Tree

   For a multicast tree, each node needs a tree label to identify a
   multicast tree.  There are three options for tree label assignment:

   o  From each router's SRLB that the controller learns

   o  From the common SRGB that the controller learns

   o  From the controller's local label space

The detailed instruction is as per
[I-D.ietf-bess-bgp-multicast-controller].  Section 3.5.1 defines a
new TLV to extend the CCI Object-Type support the tree label.

The operations for initiating multicast tree LSPs and
downloadinglabels are consistent with [I-D.ietf-pce-sr-p2mp-policy].

### 2.4.2.  SR-P2MP Tree

The processing of PCE-based SR P2MP policy delivery and P2MP path
establishment has been defined in [I-D.ietf-pce-sr-p2mp-policy].  For
the calculated multicast tree, SR P2MP multicast tree update can
refer to that document.

### 2.4.3.  BIER Tree

Different from other forwarding technologies, BIER does not need the
transit nodes to maintain the multicast state, and only needs to
forward and encapsulate the packets according to the BitString and
BIFT.  BIFT information can be learned by IGP.  For each node of a
sub-domain, the BIFT can be configured locally or allocated by BGP
controller or PCE.  Allocation by PCE is as per
[I-D.chen-pce-pcep-extension-pce-controller-bier].

After receiving the BFR information reported from LHRs, the
controller combine BFR information of LHRs into a BitString to guide
multicast data forwarding.

The controller (PCE) sends the BitString to the FHR via PCInitiate or
PCUpd message carrying Tree Forwarding State Synchronization (TFSS)
object defined in Section 3.6 to inform the FHR to forward multicast
data for a specific multicast tree according to the BitString, and
the tree identifier is (*, g) or (s, g) tuple.

### 2.5.  Multicast Statistics

Multicast statistics are helpful for multicast service analysis and
business development.  This section describes how to collect
statistics about multicast users.

For each LHR, the statistics consist of two parts, one is the number
of local users in the its managed domain, and the other is the number
of other managed domains.  The sum of these two parts is the number
of multicast data that the one LHR needs to duplicate.

LHRs send this information to the controller via PCRpt message
carrying Multicast Receiver Statistics (MRS) object.  LHRs need
report this information to the controller periodically.  Once the

statistics information of a LHR change, the LHR also needs reports to the controller.

Controller collects the statistics reported by LHRs and MAY periodically informs the FHR of the statistics via PCRpt messages carrying MRS object so that the FHR can feed the statistics back to the multicast source.

## 3.  PCEP Object Formats

### 3.1.  OPEN Object

#### 3.1.1.  STATEFUL-PCE-CAPABILITY TLV

During the PCEP initialization phase, PCEP speakers advertise stateful capability via the STATEFUL-PCE-CAPABILITY TLV in the OPEN object.  Various flags are defined for the STATEFUL-PCE-CAPABILITY TLV defined in [RFC8231] and updated in [RFC8232] and [RFC8281].

A new flag is added in this document, whose code point is TBD1:

MULTICAST-STATE-CAPABILITY bit, if this bit is set to 1 by a PCEP speaker, it indicates that the PCEP speaker supports the capability of these new extensions as specified in this document.

To support the multicast tree state management capabilities described in this document, both PCE and PCC need to set MULTICAST-STATE-CAPABILITY bit.  If a PCEP speaker receivers PECP message with the newly defined object, but without the MULTICAST-STATE-CAPABILITY bit set in STATEFUL-PCE-CAPABILITY TLV in the OPEN object, it MUST:

o  Send a PCErr message with Error-Type=10(Reception of an invalid object) and Error-Value TBD2(MULTICAST-STATE-CAPABILITY bit is not set).

o  Terminate the PCEP session.

#### 3.1.2.  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 [RFC9050].

[I-D.dhody-pce-pcep-extension-pce-controller-p2mp] adds a new flag (M Bit) in the PCECC-CAPABILITY sub-TLV to indicate the support for P2MP in PCECC, which is applicable for multicast tree described in this document as well.

## 3.2.  PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; [RFC9050] defines a
PST value for PCECC as '2', which is applicable for multicast tree
described in this document as well.

## 3.3.  Multicast Source Registration Object

The MSR object is optional and SHOULD contain multicast source
information and FHR information.  The MSR object SHOULD be carried
within a PCRpt message sent by PCC to PCE for registration.  The MSR
object SHOULD be carried within a PCUpd message sent by PCE to PCC in
response to registration.

MSR Object-Class is TBD3.  MSR Object-Type is 1.  The format of the
MSR object body is:
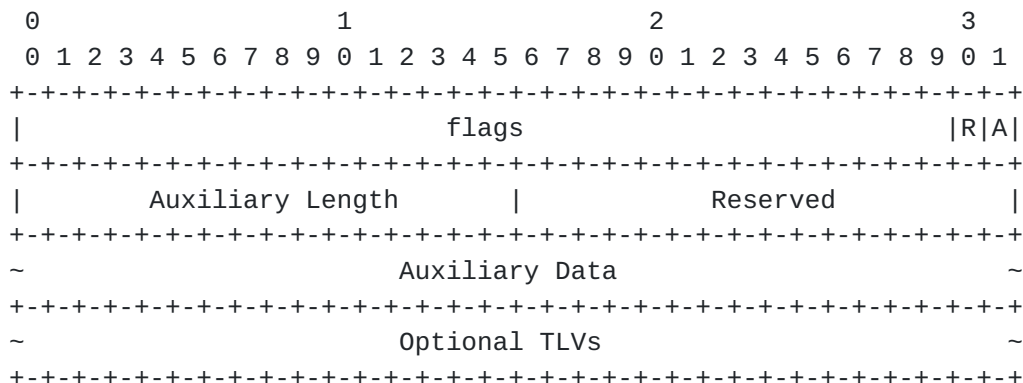
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         flags                          |R|A|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Auxiliary Length      |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                       Auxiliary Data                        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                       Optional TLVs                         ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
                   Figure 1: MSR Object Body Format

R (Register flag, 1 bit): The R flag set to 1 indicates that the PCC
is registering multicast information to the PCE.  The R flag set to 0
indicates that the PCC revokes the registration.

A (Authentication flag, 1 bit): The A flag set to 1 indicates success
of registration.  The A flag set to 0 indicates failure of
registration or revocation of registration.  If there is no
authentication information, A flag SHOULD also be set to 0.

Auxiliary Length (1 Octet): indicates the length of Auxiliary Data.

Auxiliary Data (Variable length): contains functional data such as
authentication information.

MSR object MAY include Multicast Address TLV, mLDP FEC TLV, SR-IPV4-
P2MP-POLICY-ID TLV, SR-IPV6-P2MP-POLICY-ID TLV, and VPN Information
TLV.  SR-IPV4-P2MP-POLICY-ID TLV and SR-IPV6-P2MP-POLICY-ID TLV are

defined in [I-D.ietf-pce-sr-p2mp-policy].  MSR object carries
different TLVs depending on the multicast tree.

### 3.3.1.  Multicast Address TLV

In IP multicast scenarios, Multicast Address TLV SHOULD be included.

The format of the Multicast Address TLV is:
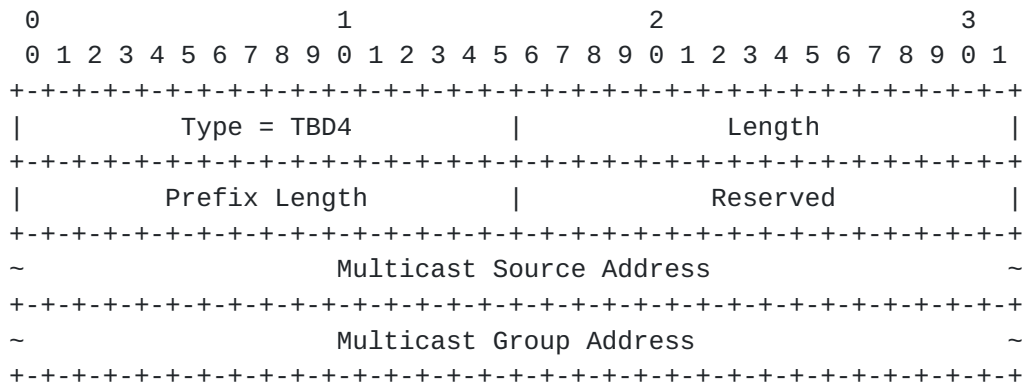
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type = TBD4           |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Prefix Length          |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     Multicast Source Address                  ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     Multicast Group Address                   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
                  Figure 2: Multicast Address TLV Format

Prefix Length (2 Octets): indicates the length of multicast
addresses. the length MAY be 4 Octets, 8 Octets, 16 Octets, or 32
Octets.  The multicast source address can be empty, but the multicast
group address must be filled.  If both the multicast source address
and multicast group address exist, they must be both IPv4 and IPv6
addresses.

Multicast Source Address (Variable length): contains IPv4 or IPv6
address of the multicast source.

Multicast Group Address (Variable length): contains IPv4 or IPv6
address of the multicast group.

### 3.3.2.  mLDP FEC TLV

In mLDP multicast scenarios, Multicast Address TLV SHOULD be
included.

The format of the mLDP FEC TLV is:

```
 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 = TBD5          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                            mLDP FEC                           ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                  Figure 3: mLDP FEC TLV Format
```
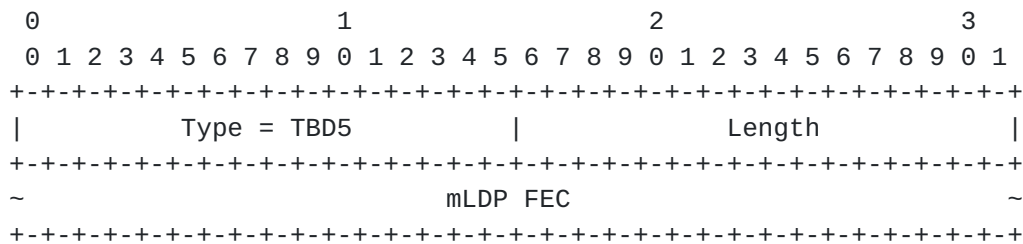
   mLDP FEC (Variable length): carries mLDP FEC.

### 3.3.3.  VPN Information TLV

   VPN Information TLV is used to report VPN information about multicast
   sources and receivers.  The format of the VPN Information TLV is:

```
 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 = TBD6          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         VPN Identifier                        |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                 Figure 4: VPN Information TLV Format
```
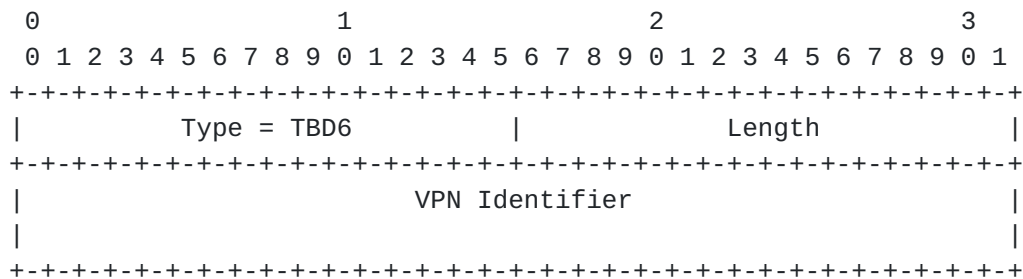
   VPN Identifier (8 Octets): indicates the VPN which the receiver used.

### 3.4.  Multicast Receiver Information Object

   The MRI object is optional and SHOULD contain receivers' information
   for matching the multicast registration information.  The MRI object
   SHOULD be carried within a PCRpt message sent by PCC to PCE for
   multicast joining or leaving.

   MRI Object-Class is TBD7.  MRI Object-Type is 1.  The format of the
   MRI object body is:

```
 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       |B|S|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                          Optional TLVs                        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                  Figure 5: MRI Object Body Format
```

B(BIER multicast flag, 1 bit): The B flag set to 1 indicates that
multicast protocol is BIER.  If the B flag set to 1, MRI object
SHOULD include BFR information TLV defined in Section 3.4.1.

S(Subscribe flag, 1 bit): The S flag set to 1 indicates that PCC
delivers the message requesting to join PCE.  The S flag set to 0
indicates that PCC delivers the message requesting to leave to PCE.

MRI object MAY include Multicast Address TLV, mLDP FEC TLV, SR-IPV4-
P2MP-POLICY-ID TLV, SR-IPV6-P2MP-POLICY-ID TLV, VPN Information TLV,
and BFR information TLV.

## 3.4.1.  BFR Information TLV

BFR Information TLV is used to report router location information in
the BIER domain.  The format of the BFR Information TLV is:

```
 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 = TBD8          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Subdomain-Id  |            BFR-ID              | BSL  |Padding|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: BFR Information TLV Format

Subdomain-Id (1 Octet): Unique value identifying the BIER subdomain.

BFR-ID (2 Octets): Identification of BFR in a subdomain.

BSL (BitString Length, 4 bits): encodes the length in bits of the
BitString as per [RFC8296], the maximum length of the BitString is 7,
it indicates the length of BitString is 4096.  It is used to refer to
the number of bits in the BitString.

## 3.5.  CCI Object

The CCI object [RFC9050] is used by the PCE to specify the forwarding
instructions to the PCC and MAY be carried within a PCInitiate or
PCRpt message for control information download/report.

The CCI Object Type 1 for MPLS Label is defined in [RFC9050], which
is used for the P2MP LSPs as well.  For mLDP multicast trees, in
addition to forwarding labels, corresponding tree labels or label
stacks are required.  Therefore, a new TLV is needed to carry tree
labels or label stacks.

### 3.5.1.  MPLS Tree Label TLV

The format of MPLS Tree Label TLV is:
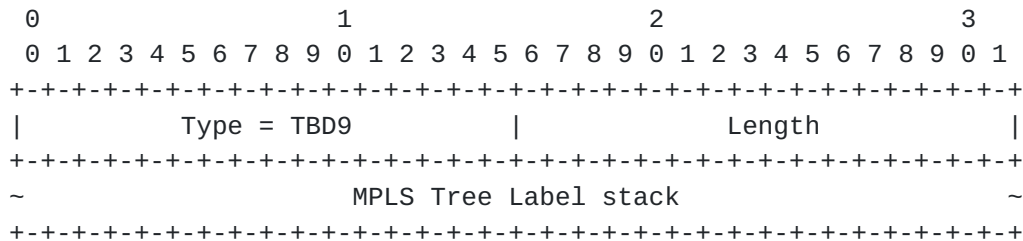
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type = TBD9          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     MPLS Tree Label stack                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
              Figure 7: MPLS Tree Label TLV Format

MPLS Tree Label stack (Variable length): contains tree labels used to
identify multicast trees.  There are three application scenarios, as
per [I-D.ietf-bess-bgp-multicast-controller].

### 3.6.  Tree Forwarding State Synchronization Object

The TFSS object is optional and SHOULD contain the forwarding state
of control plane that the tree node needs to synchronize.  The TFSS
object SHOULD be carried within a PCInitiate or PCUpd message sent by
PCE to PCC, or a PCRpt message sent by PCC to PCE for response.

TFSS Object-Class is TBD10.  TFSS Object-Type is 1.  The format of
the TFSS object body is:
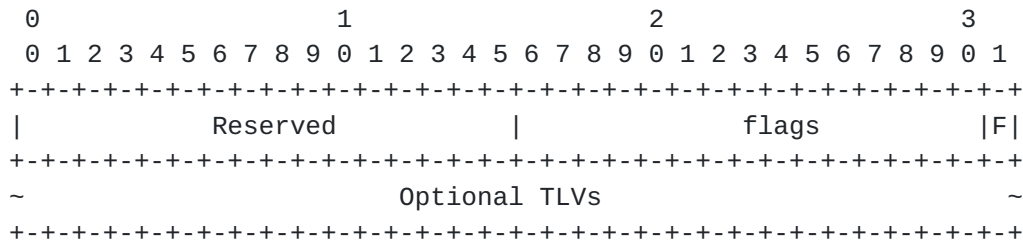
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Reserved           |            flags           |F|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                        Optional TLVs                          ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
                Figure 8: TFSS Object Body Format

F (Forwarding flag, 1 bit): The F flag set to 1 indicates that the
node MAY start forwarding multicast packets.  The F flag set to 0
indicates that the node SHOULD stop forwarding multicast packets.

TFSS object MAY include Multicast Address TLV, VPN Information TLV,
and BIER Attribute TLV.  When TT=1, BIER Attribute TLV SHOULD be
included.

### 3.6.1.  BIER Attribute TLV

BIER Attribute TLV is used to instruct the FHR to forward multicast
packets to the users according to the BitString specified by the
controller.  The format of BIER Attribute TLV is:

```
 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            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Subdomain-id  |      SI       | BSL  ~       BitString        ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
                Figure 9: BIER Attribute TLV Format

Subdomain-id (1 Octet): Unique value identifying the BIER subdomain.

SI (Set Identifier, 1 Octet): encoding the Set Identifier used in the
encapsulation for this BIER subdomain for this BitString length.

BSL (BitString Length, 4 bits): encodes the length in bits of the
BitString as per [RFC8296], the maximum length of the BitString is 7,
it indicates the length of BitString is 4096.  It is used to refer to
the number of bits in the BitString.

BitString(Variable length): indicates the path of multicast data
packets forwarding for headend.

### 3.7.  Multicast Receiver Statistics Object

The MRS object is optional and used to inform PCE of the number of
receivers.  The MRS object SHOULD be carried within a PCRpt or a
PCUpd message for synchronize receiver information periodically, or
PCRpt message for the leaving of receivers.

MRS Object-Class is TBD12.  MRS Object-Type is 1.  The format of the
MRS object body is:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Number Length        |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Multicast Statistics                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                        Optional TLVs                          ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
                Figure 10: MRS Object Body Format

Number Length (2 Octets): indicates the length of receiver number.

Multicast Statistics (4 Octets): indicates the statistics information
for a particular multicast tree of a controller or a LHR.

MRS object MAY include Multicast Address TLV, mLDP FEC TLV, SR-IPV4-
P2MP-POLICY-ID TLV, and SR-IPV6-P2MP-POLICY-ID TLV.

## 4.  Message Formats

### 4.1.  PCRpt message

The definition of the PCRpt message from [RFC8231] and [RFC9050] is
extended to optionally include MSR object, MRI object, TFSS object,
and MRS object after the path object.  The encoding will become:

```
    <PCRpt Message> ::= <Common Header>
                        <state-report-list>
Where:

    <state-report-list> ::= <state-report>[<state-report-list>]

    <state-report> ::= [<SRP>]
                        <LSP>
                        <path>
                        [<MSR>]
                        [<MRI>]
                        [<TFSS>]
                        [<MRS>]


Where:
  <path> is as per [RFC8231] and the LSP and SRP object are
  also defined in [RFC8231].
```

### 4.2.  PCInitiate message

The definition of the PCInitiate message from [RFC8281] is extended
to optionally include TFSS object after the path object.  The
encoding will become:

```
   <PCUpd Message> ::= <Common Header>
                         <update-request-list>
   Where:

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

      <update-request> ::= <SRP>
                            <LSP>
                            [<TFSS>]

   Where:
      LSP and SRP object are defined in [RFC8231].
```

## 4.3.  PCUpd message

The definition of the PCUpd message from [RFC8231] is extended to
optionally include MSR object, TFSS object and MRS object after the
path object.  The encoding will become:

```
      <PCUpd Message> ::= <Common Header>
                            <update-request-list>
   Where:

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

      <update-request> ::= <SRP>
                            <LSP>
                            <path>
                            [<MSR>]
                            [<TFSS>]
                            [<MRS>]

   Where:
      <path> is as per [RFC8231] and the LSP and SRP object are
      also defined in [RFC8231].
```
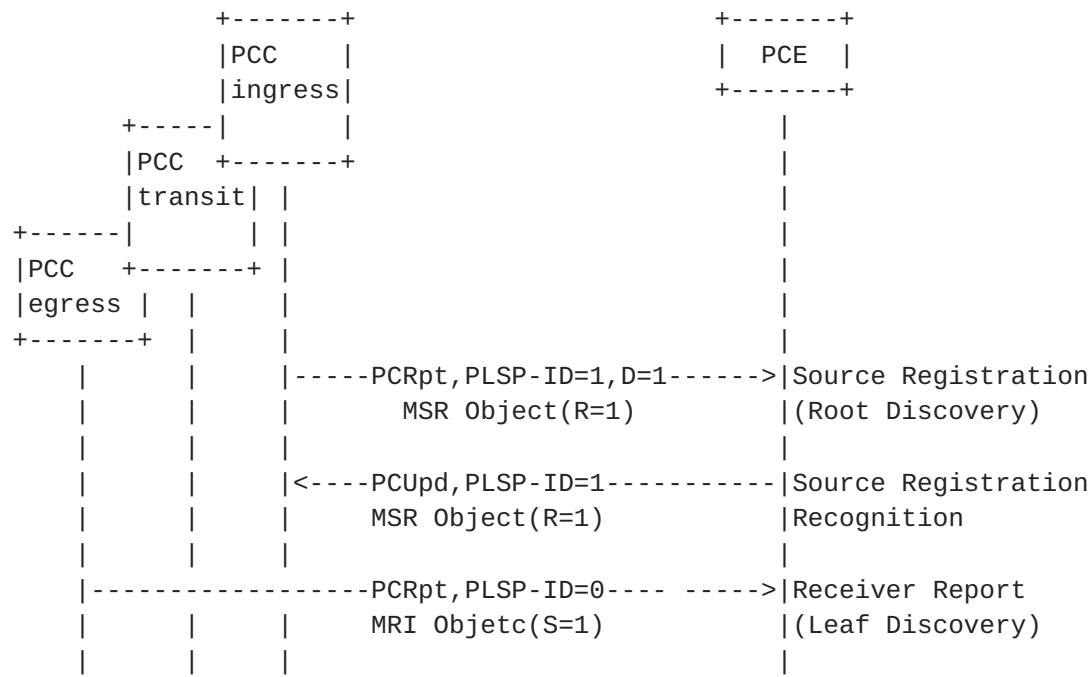
## 5.  Example Workflow

## 5.1.  Multicast Tree Discovery

```
                   +-------+                    +-------+
                   |PCC    |                    |  PCE  |
                   |ingress|                    +-------+
           +-----|         |                        |
           |PCC  +-------+                           |
           |transit| |                              |
      +------|      | |                             |
      |PCC   +-------+ |                             |
      |egress |  |     |                            |
      +-------+  |     |                            |
          |      |     |-----PCRpt,PLSP-ID=1,D=1------>|Source Registration
          |      |     |        MSR Object(R=1)        |(Root Discovery)
          |      |     |                              |
          |      |     |<----PCUpd,PLSP-ID=1-----------|Source Registration
          |      |     |      MSR Object(R=1)          |Recognition
          |      |     |                              |
          |----------------PCRpt,PLSP-ID=0---- ----->|Receiver Report
          |      |     |      MRI Objetc(S=1)          |(Leaf Discovery)
          |      |     |                              |
           Figure 11: Workflow of Multicast Tree Discovery
```

## 5.2.  Multicast Tree Path Establishment and Update

## 5.2.1.  mLDP Tree and SR-P2MP Tree

   The workflow of mLDP Tree and SR-P2MP Tree are as per
   [I-D.ietf-pce-sr-p2mp-policy].

## 5.2.2.  BIER Tree

```
                  +-------+                      +-------+
                  |PCC    |                      |  PCE  |
                  |ingress|                      +-------+
            +------|       |                          |
            |      +-------+                          |
            | transit| |                             |
       +------|       | |                             |
       |      +--------+ |                            |
       |egress |  |      |                            |
       +--------+ |      |                            |
            |     |      |<-----PCInitiate,PLSP-ID=1-------|Tree State
            |     |      |      TFSS Objetc (F=1,TT=1)     |Setup
            |     |      |                                 |
            |     |      |------PCRpt,PLSP-ID=1----------->|Tree State
            |     |      |                                 |Recognition
            |     |      |                                 |
            |------------------PCRpt,PLSP-ID=0----------->|Receiver Report
            |     |      |      MRI Objetc(S=1/S=0)         |(Leaf Discovery)
            |     |      |                                 |
            |     |      |<-----PCUpd,PLSP-ID=1------------|Tree State
            |     |      |      TFSS Objetc (F=1,TT=1)     |Update
            |     |      |                                 |
            |     |      |------PCRpt,PLSP-ID=1----------->|Tree State
            |     |      |                                 |Recognition
            |     |      |                                 |
                 Figure 12: Workflow of Multicast Tree Discovery
```
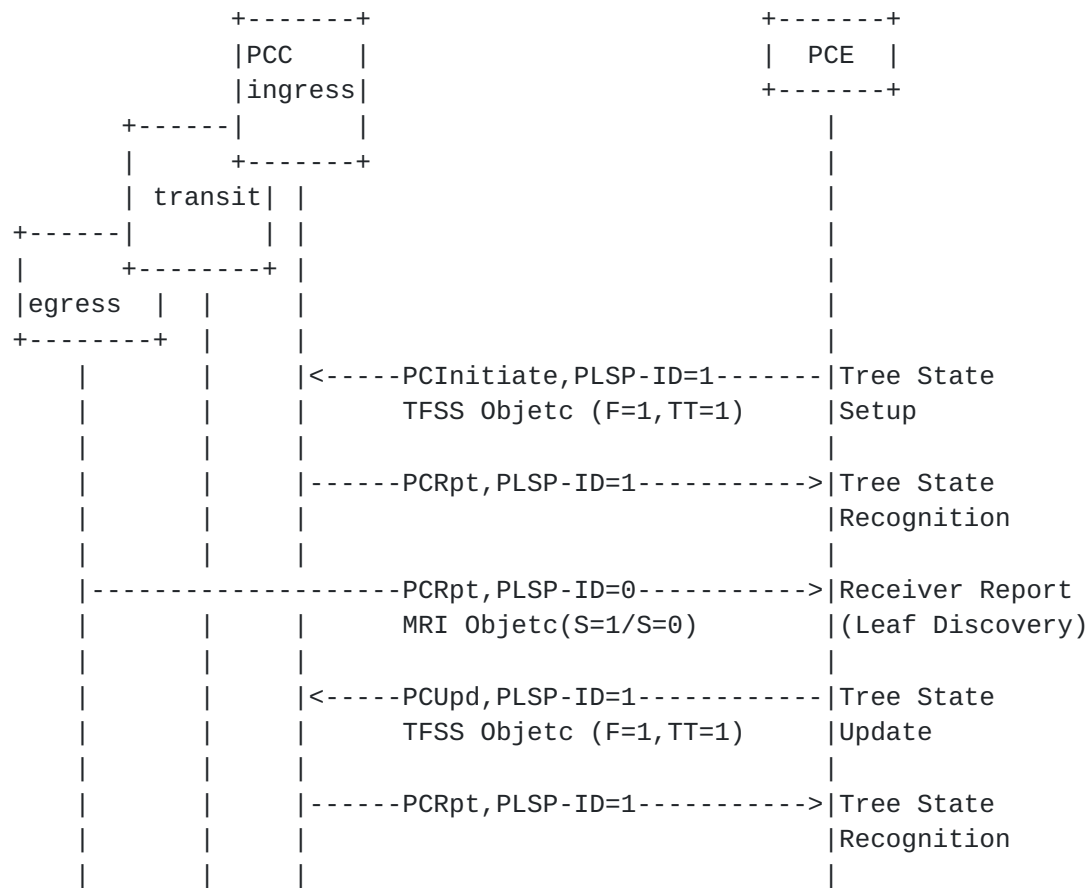
   In the previous multicast tree discovery process, ingress has
   delegated the PCE to calculate the multicast path.  Therefore, during
   the establishment of the multicast tree, PCE informs the ingress to
   forward multicast data for (S,G)/(*,G) via PCInitiate message
   carrying TFSS object according to the delegation.  The PLSP-ID is the
   value specified by the ingress.

   Ingress then updates the tree state and responds to the PCE.  At this
   point, the BIER tree is formally established and the ingress starts
   forwarding for the multicast according to the BitString specified by
   PCE.

   If the topology of the BIER tree changes, the corresponding egresses
   will send PCRpt messages carring MRI object to PCE to inform PCE of
   their changes.

   The PCE updates the BitStrig based on changes of the egresses and
   informs the ingress to update.  State update process is similar to
   state setup process, except for the type of message sent by the PCE.

## [6](#). Security Considerations

   To be added.

## [7](#).  IANA Considerations

### [7.1](#).  MULTICAST-STATE-CAPABILITY

   IANA is requested to allocate a new code point within registry
   "STATEFUL-PCE-CAPABILITY TLV Flag Field" under "Path Computation
   Element Protocol (PCEP) Numbers" as follows:

| Value | Description | Reference |
| ------------ | ------------------------------ | --------------- |
| TBD1 | MULTICAST-STATE-CAPABILITY | This document |

### [7.2](#).  PCEP-ERROR Object

   IANA is requested to allocate code-points in the "PCEP-ERROR Object
   Error Types and Values" sub-registry for the following new error-type
   and error-value:

| Error-Type | Description | Reference |
| ------------ | ------------------------------ | --------------- |
| 10 | Error-value = TBD2  MULTICAST-STATE-CAPABILITY  bit is not set | This document |

### [7.3](#).  New Objects

   IANA is requested to allocate the following Object-Class Values in
   the "PCEP Objects" sub-registry under the "Path Computation Element
   Protocol (PCEP) Numbers" registry:

| Object-Class Value | Description | Reference |
| ------------------ | ------------------------------ | -------------- |
| TBD3 | Multicast Source Registration | This document |
| TBD7 | Multicast Receiver Information | This document |
| TBD10 | Tree Forwarding State Synchronization | This document |
| TBD12 | Multicast Receiver Statistics | This document |

   IANA is requested to allocate the following Object-Type Value of CCI
   object in the "PCEP Objects" sub-registry under the "Path Computation
   Element Protocol (PCEP) Numbers" registry:

### 7.4.  New TLVs

   IANA is requested to allocate the following TLVs:

| Type | Description | Reference |
|------|-------------|-----------|
| TBD4 | Multicast Source Address | This document |
| TBD5 | mLDP FEC | This document |
| TBD6 | VPN Information | This document |
| TBD8 | BFR Information | This document |
| TBD9 | MPLS Tree Label | This document |
| TBD10 | BIER Attribute | This document |

### 8.  Acknowledgements

   The author thanks ... for their review and comments.

### 9.  References

### 9.1.  Normative References

   [I-D.ietf-pce-sr-p2mp-policy]
             Bidgoli, H., Voyer, D., Rajarathinam, S., Hemmati, E.,
             Saad, T., and S. Sivabalan, "PCEP extensions for p2mp sr
             policy", draft-ietf-pce-sr-p2mp-policy-00 (work in
             progress), December 2021.

   [RFC1112]  Deering, S., "Host extensions for IP multicasting", STD 5,
             RFC 1112, DOI 10.17487/RFC1112, August 1989,
             <https://www.rfc-editor.org/info/rfc1112>.

   [RFC2362]  Estrin, D., Farinacci, D., Helmy, A., Thaler, D., Deering,
             S., Handley, M., Jacobson, V., Liu, C., Sharma, P., and L.
             Wei, "Protocol Independent Multicast-Sparse Mode (PIM-SM):
             Protocol Specification", RFC 2362, DOI 10.17487/RFC2362,
             June 1998, <https://www.rfc-editor.org/info/rfc2362>.

   [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>.

   [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>.

   [RFC5015]  Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano,
              "Bidirectional Protocol Independent Multicast (BIDIR-
              PIM)", RFC 5015, DOI 10.17487/RFC5015, October 2007,
              <https://www.rfc-editor.org/info/rfc5015>.

   [RFC6388]  Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B.
              Thomas, "Label Distribution Protocol Extensions for Point-
              to-Multipoint and Multipoint-to-Multipoint Label Switched
              Paths", RFC 6388, DOI 10.17487/RFC6388, November 2011,
              <https://www.rfc-editor.org/info/rfc6388>.

   [RFC6514]  Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP
              Encodings and Procedures for Multicast in MPLS/BGP IP
              VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012,
              <https://www.rfc-editor.org/info/rfc6514>.

   [RFC7761]  Fenner, B., Handley, M., Holbrook, H., Kouvelas, I.,
              Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent
              Multicast - Sparse Mode (PIM-SM): Protocol Specification
              (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March
              2016, <https://www.rfc-editor.org/info/rfc7761>.

   [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>.

   [RFC8279]  Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A.,
              Przygienda, T., and S. Aldrin, "Multicast Using Bit Index
              Explicit Replication (BIER)", RFC 8279,
              DOI 10.17487/RFC8279, November 2017,
              <https://www.rfc-editor.org/info/rfc8279>.

   [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>.

   [RFC8296]  Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A.,
              Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation
              for Bit Index Explicit Replication (BIER) in MPLS and Non-
              MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January
              2018, <https://www.rfc-editor.org/info/rfc8296>.

   [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>.

   [RFC9050]  Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "Path
              Computation Element Communication Protocol (PCEP)
              Procedures and Extensions for Using the PCE as a Central
              Controller (PCECC) of LSPs", RFC 9050,
              DOI 10.17487/RFC9050, July 2021,
              <https://www.rfc-editor.org/info/rfc9050>.

## 9.2.  Informative References

   [I-D.chen-pce-pcep-extension-pce-controller-bier]
              Chen, R. and B. Xu, "PCEP Procedures and Protocol
              Extensions for Using PCE as a Central Controller (PCECC)
              of BIER", draft-chen-pce-pcep-extension-pce-controller-
              bier-02 (work in progress), December 2021.

   [I-D.dhody-pce-pcep-extension-pce-controller-p2mp]
              Li, Z., Peng, S., Geng, X., and M. S. Negi, "Path
              Computation Element Communication Protocol (PCEP)
              Procedures and Extensions for Using the PCE as a Central
              Controller (PCECC) of point-to-multipoint (P2MP) LSPs",
              draft-dhody-pce-pcep-extension-pce-controller-p2mp-08
              (work in progress), March 2022.

   [I-D.ietf-bess-bgp-multicast-controller]
              Zhang, Z., Raszuk, R., Pacella, D., and A. Gulko,
              "Controller Based BGP Multicast Signaling", draft-ietf-
              bess-bgp-multicast-controller-07 (work in progress), July
              2021.

   [I-D.ietf-pim-sr-p2mp-policy]
              (editor), D. V., Filsfils, C., Parekh, R., Bidgoli, H.,
              and Z. Zhang, "Segment Routing Point-to-Multipoint
              Policy", draft-ietf-pim-sr-p2mp-policy-03 (work in
              progress), August 2021.

[I-D.ietf-spring-sr-replication-segment]
          (editor), D. V., Filsfils, C., Parekh, R., Bidgoli, H.,
          and Z. Zhang, "SR Replication Segment for Multi-point
          Service Delivery", draft-ietf-spring-sr-replication-
          segment-06 (work in progress), October 2021.

[I-D.zzhang-mvpn-evpn-controller]
          Zhang, Z., Parekh, R., Zhang, Z., and H. Bidgoli, "MVPN
          and EVPN BUM Signaling with Controllers", draft-zzhang-
          mvpn-evpn-controller-01 (work in progress), October 2021.

Authors' Addresses

   Huanan Li
   China Telecom

   Email: lihn6@foxmail.com


   Aijun Wang
   China Telecom

   Email: wangaj3@chinatelecom.cn


   Zhaohui Zhang
   Juniper Networks

   Email: zzhang@juniper.net


   Huaimo Chen
   Futurewei

   Email: Huaimo.chen@futurewei.com


   Ran Chen
   ZTE Corporation

   Email: chen.ran@zte.com.cn