PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 11, 2021

A. Wang
China Telecom
B. Khasanov
Yandex LLC
S. Fang
R. Tan
Huawei Technologies,Co.,Ltd
C. Zhu
ZTE Corporation

February 7, 2021

PCEP Extension for Native IP Network
draft-ietf-pce-pcep-extension-native-ip-11

Abstract

This document defines the Path Computation Element Communication Protocol (PCEP) extension for Central Control Dynamic Routing (CCDR) based application in Native IP network. The scenario and framework of CCDR in native IP is described in [RFC8735] and [I-D.ietf-teas-pce-native-ip]. This draft describes the key information that is transferred between Path Computation Element (PCE) and Path Computation Clients (PCC) to accomplish the End to End (E2E) traffic assurance in Native IP network under central control mode.

Status of This Memo

Table of Contents

1. Introduction

Generally, Multiprotocol Label Switching Traffic Engineering (MPLS-TE) requires the corresponding network devices support Multiprotocol Label Switching (MPLS) or Resource ReSerVation Protocol (RSVP)/Label Distribution Protocol (LDP) technologies to assure the End-to-End (E2E) traffic performance. But in native IP network, there will be no such signaling protocol to synchronize the action among different network devices. It is necessary to use the central control mode that described in [RFC8283] to correlate the forwarding behavior among different network devices. Draft [I-D.ietf-teas-pce-native-ip] describes the architecture and solution philosophy for the E2E traffic assurance in Native IP network via Multi Border Gateway Protocol (BGP) solution. This draft describes the corresponding Path Computation Element Communication Protocol (PCEP) extensions to transfer the key information about BGP peer info, peer prefix association and the explicit peer route on on-path routers.

2. Conventions used in this document

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

This document uses the following terms defined in [RFC5440]: PCE, PCEP

The following terms are defined in this document:

- CCDR: Central Control Dynamic Routing
- E2E: End to End
- BPI: BGP Peer Info
- EPR: Explicit Peer Route
- PPA: Peer Prefix Association
- QoS: Quality of Service

4. Capability Advertisemnt

4.1. Open message

During the PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of Native IP extensions.

This document defines a new Path Setup Type (PST) [RFC8408] for Native-IP, as follows:

- o PST = TBD1: Path is a Native IP path as per [I-D.ietf-teas-pce-native-ip].

A PCEP speaker MUST 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.

[I-D.ietf-pce-pcep-extension-for-pce-controller] defined the PCECC-CAPABILITY sub-TLV to exchange information about their PCECC capability. A new flag is defined in PCECC-CAPABILITY sub-TLV for Native IP.

N (NATIVE-IP-TE-CAPABILITY - 1 bit - TBD2): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable for TE in Native IP network as specified in this document. The flag MUST be set by both the PCC and PCE in order to support this extension.

If a PCEP speaker receives the PATH-SETUP-TYPE-CAPABILITY TLV with the newly defined path setup type, but without the N bit set in PCECC-CAPABILITY sub-TLV, it MUST:

- o Send a PCErr message with Error-Type=10(Reception of an invalid object) and Error-Value TBD3(PCECC NATIVE-IP-TE-CAPABILITY bit is not set).

- o Terminate the PCEP session

5. PCEP messages

PCECC Native IP TE solution utilizing the existing PCE LSP Initate Request message(PCInitiate)[RFC8281], and PCE Report message(PCRpt) [RFC8281] to accomplish the multi BGP sessions establishment, E2E TE path deployment, and route prefixes advertisement among different BGP sessions. A new PST for Native-IP is used to indicate the path setup based on TE in Native IP networks.

The extended PCInitiate message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is used to download

or cleanup central controller's instructions (CCIs). [I-D.ietf-pce-pcep-extension-for-pce-controller] specify an object called CCI for the encoding of central controller's instructions. This document specify a new CCI object-type for Native IP. The PCEP messages are extended in this document to handle the PCECC operations for Native IP. Three new PCEP Objects (BGP Peer Info (BPI) Object, Explicit Peer Route (EPR) Object and Peer Prefix Association (PPA) Object) are defined in this document. Refer toSection 7 for detail object definitions.

5.1. The PCInitiate message

The PCInitiate Message defined in [RFC8281] and extended in [I-D.ietf-pce-pcep-extension-for-pce-controller] is further extended to support Native-IP CCI.

The format of the extended PCInitiate message is as follows:

```
     <PCInitiate Message> ::= <Common Header>
                              <PCE-initiated-lsp-list>
   Where:
     <Common Header> is defined in [RFC5440]

     <PCE-initiated-lsp-list> ::= <PCE-initiated-lsp-request>
                                  [<PCE-initiated-lsp-list>]

     <PCE-initiated-lsp-request> ::=
                          (<PCE-initiated-lsp-instantiation>|
                           <PCE-initiated-lsp-deletion>|
                           <PCE-initiated-lsp-central-control>)

     <PCE-initiated-lsp-central-control> ::= <SRP>
                                             (<LSP>
                                             <cci-list>)|
                                             ((<BPI>|<EPR>|<PPA>)
                                             <CCI>)

     <cci-list> ::=  <CCI>
                     [<cci-list>]

   Where:
      <cci-list> is as per
      [I-D.ietf-pce-pcep-extension-for-pce-controller].
      <PCE-initiated-lsp-instantiation> and
      <PCE-initiated-lsp-deletion> are as per
      [RFC8281].
```

The LSP and SRP object is defined in [RFC8231].

When PCInitiate message is used create Native IP instructions, the SRP and CCI objects MUST be present. The error handling for missing SRP or CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. Further only one of BPI, EPR, or PPA object MUST be present. The PLSP-ID within the LSP object should be set by PCC uniquely according to the Symbolic Path Name TLV that included in the CCI object. The Symbolic Path Name is used by the PCE/PCC to identify uniquely the E2E native IP TE path.

If none of them are present, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD4 (Native IP object missing). If there are more than one of BPI, EPR or PPA object are presented, the receiving PCC MUST send a PCErr message with Error-type=19(Invalid Operation) and Error-value=TBD5(Only one of the BPI, EPR or PPA object can be included in this message).

To cleanup the SRP object must set the R (remove) bit.

5.2. The PCRpt message

The PCRpt message is used to acknowledge the Native-IP instructions received from the central controller (PCE).

The format of the PCRpt message is as follows:

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

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

      <state-report> ::= (<lsp-state-report>|
                          <central-control-report>)

      <lsp-state-report> ::= [<SRP>]
                             <LSP>
                             <path>

      <central-control-report> ::= [<SRP>]
                                   (<LSP>
                                   <cci-list>)|
                                   ((<BPI>|<EPR>|<PPA>)
                                   <CCI>)


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

The error handling for missing CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. Further only one of BPI, EPR, or PPA object MUST be present.

If none of them are present, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD4 (Native IP object missing). If there are more than one of BPI, EPR or PPA object are presented, the receiving PCE MUST send a PCErr message with Error-type=19(Invalid Operation) and Error-value=TBD5(Only one of the BPI, EPR or PPA object can be included in this message).

6. PCECC Native IP TE Procedures

The detail procedures for the TE in native IP environment are described in the following sections.

6.1. BGP Session Establishment Procedures

The procedures for establishing the BGP session between two peers is shown below, using the PCInitiate and PCRpt message pair.

The PCInitiate message should be sent to PCC which acts as BGP routers and route reflector. In the example in Figure 1, it should be sent to R1(M1), R3(M2 & M3) and R7(M4), when R3 acts as RR.

When PCC receives the BPI and CCI object (with the R bit set to 0 in SRP object) in PCInitiate message, the PCC should try to establish the BGP session with the indicated Peer AS and Local/Peer IP address.

When PCC creates successfully the BGP session that is indicated by the associated information, it should report the result via the PCRpt messages, with BPI object included, and the corresponding SRP and CCI object.

When PCC receives this message with the R bit set to 1 in SRP object in PCInitiate message, the PCC should clear the BGP session that indicated by the BPI object.

When PCC clears successfully the specified BGP session, it should report the result via the PCRpt message, with the BPI object included, and the corresponding SRP and CCI object.

```
                    +------------------+
        +-----------+       PCE        +----------+
        |           +--------^---------+          |
        |                    |                    |
                    M2/M2-R & M3/M3-R
        |                    |                    |
        |               +---v---+                 |
        +---------------+ R3(RR)+-----------------+
        |               +-------+                 |
     M1/M1-R                                   M4/M4-R
        |                                         |
       +v-+          +--+          +--+         +-v+
       |R1+----------+R5+----------+R6+---------+R7|
       ++-+          +--+          +--+         +-++
        |                                         |
        |            +--+          +--+           |
        +------------+R2+----------+R4+-----------+
                     +--+          +--+
```

Figure 1: BGP Session Establishment Procedures(R3 act as RR)

The message number, message peers, message type and message key parameters in the above figures are shown in below table:

Table 1: Message Information

No.	Peers	Type	Message Key Parameters
M1	PCE/R1	PCInitiate	CC-ID=X1(Symbolic Path Name=Class A)
M1-R		PCRpt	BPI Object(Local_IP=R1_A,Peer_IP=R3_A)
M2	PCE/R3	PCInitiate	CC-ID=X2(Symbolic Path Name=Class A)
M2-R		PCRpt	BPI Object(Local_IP=R3_A,Peer_IP=R1_A)
M3	PCE/R3	PCInitiate	CC-ID=X3(Symbolic Path Name=Class A)
M3-R		PCRpt	BPI Object(Local_IP=R3_A,Peer_IP=R7_A)
M4	PCE/R7	PCInitiate	CC-ID=X4(Symbolic Path Name=Class A)
M4-R		PCRpt	BPI Object(Local_IP=R7_A,Peer_IP=R3_A)

If the PCC cannot establish the BGP session that required by this object, it should report the error values via PCErr message with the newly defined error type(Error-type=TBD6) and error value(Error-value=TBD7, Peer AS not match; or Error-Value=TBD8, Peer IP can't be reached), which is indicated in Section 9

If the Local_IP or Peer_IP within BPI object is used in other existing BGP sessions, the PCC should report such error situation via PCErr message with Err-type=TBD6 and error value(Error-value=TBD9, Local IP is in use; Error-value=TBD10, Remote IP is in use).

6.2. Explicit Route Establish Procedures

The detail procedures for the explicit route establishment procedures is shown below, using PCInitiate and PCRpt message pair.

The PCInitiate message should be sent to the on-path routers respectively. In the example, for explicit route from R1 to R7, the PCInitiate message should be sent to R1(M1), R2(M2) and R4(M3), as shown in Figure 2. For explicit route from R7 to R1, the PCInitiate message should be sent to R7(M1), R4(M2) and R2(M3), as shown in Figure 3..

When PCC receives the EPR and the CCI object (with the R bit set to 0 in SRP object) in PCInitiate message, the PCC should install the explicit route to the the peer.

When PCC install successfully the explicit route to the peer, it should report the result via the PCRpt messages, with EPR object included, and the corresponding SRP and CCI object.

When PCC receives the EPR and the CCI object with the R bit set to 1 in SRP object in PCInitiate message, the PCC should clear the explicit route to the peer that indicated by the EPR object.

When PCC clear successfully the explicit route that indicated by this object, it should report the result via the PCRpt message, with the EPR object included, and the corresponding SRP and CCI object.

```
               +------------------+
    +----------+       PCE        +
    |          +----^-----------^-+
    |               |           |
    |               |           |
    |               | +------+  |
    +-----------------+R3(RR)+--|-------------+
M1/M1-R             | +------+  |             |
    |               |           |             |
   +v-+      +--+   |           |   +--+    +--+
   |R1+------+R5+---+-----------|---+R6+----+R7|
   ++-+      +--+   |           |   +--+    +-++
    |            M2/M2-R      M3/M3-R         |
    |               |           |             |
    |            +--v--+     +--v-+           |
    +------------+- R2 +-----+ R4 +-----------+
                 +--+--+     +--+-+
```

Figure 2: Explicit Route Establish Procedures(From R1 to R7)

The message number, message peers, message type and message key parameters in the above figures are shown in below table:

Table 2: Message Information

No.	Peers	Type	Message Key Parameters
M1	PCE/R1	PCInitiate	CC-ID=X1(Symbolic Path Name=Class A)
M1-R		PCRpt	EPR Object(Peer Address=R7_A,Next Hop=R2_A)
M2	PCE/R2	PCInitiate	CC-ID=X2(Symbolic Path Name=Class A)
M2-R		PCRpt	EPR Object(Peer Address=R7_A,Next Hop=R4_A)
M3	PCE/R4	PCInitiate	CC-ID=X3(Symbolic Path Name=Class A)
M3-R		PCRpt	EPR Object(Peer Address=R7_A,Next Hop=R7_A)

```
                       +------------------+
                       +       PCE        +-----------+
                       +----^-----------^-+           |
                            |           |             |
                            |           |             |
                            | +------+  |             |
            +---------------+R3(RR)+--|-------------+
            |               | +------+  |         M1/M1-R
            |               |           |             |
           +--+      +--+   |           |   +--+    +-v+
           |R1+------+R5+---+-----------|---+R6+----+R7|
           ++-+      +--+   |           |   +--+    +-++
            |            M3/M3-R      M2/M2-R         |
            |               |           |             |
            |            +--v--+     +--v-+           |
            +------------+- R2 +-----+ R4 +-----------+
                         +--+--+     +--+-+
```

Figure 3: Explicit Route Establish Procedures(From R7 to R1)

The message number, message peers, message type and message key parameters in the above figures are shown in below table:

Table 3: Message Information

No.	Peers	Type	Message Key Parameters
M1	PCE/R7	PCInitiate	CC-ID=X1(Symbolic Path Name=Class A)
M1-R		PCRpt	EPR Object(Peer Address=R1_A,Next Hop=R4_A)
M2	PCE/R4	PCInitiate	CC-ID=X2(Symbolic Path Name=Class A)
M2-R		PCRpt	EPR Object(Peer Address=R1_A,Next Hop=R2_A)
M3	PCE/R2	PCInitiate	CC-ID=X3(Symbolic Path Name=Class A)
M3-R		PCRpt	EPR Object(Peer Address=R1_A,Next Hop=R1_A)

In order to avoid the transient loop during the deploy of explicit peer route, the EPR object should be sent to the PCCs in the reverse order of the E2E path. To remove the explicit peer route, the EPR object should be sent to the PCCs in the same order of E2E path.

Upon the error occurs, the PCC SHOULD send the corresponding error via PCErr message, with an error information (Error-type=TBD6, Error-value=TBD12, Explicit Peer Route Error) that defined in Section 9.

When the peer info that associated with the Symbolic Path Name is not the same as the peer info that indicated in BPI object in PCC, an

error (Error-type=TBD6, Error-value=17, EPR/BPI Peer Info mismatch) should be reported via the PCErr message.

6.3. BGP Prefix Advertisement Procedures

The detail procedures for BGP prefix advertisement is shown below, using PCInitiate and PCRpt message pair.

The PCInitiate message should be sent to PCC that acts as BGP peer router only. In the example, it should be sent to R1(M1) or R7(M2) respectively.

When PCC receives the PPA and the CCI object (with the R bit set to 0 in SRP object) in PCInitiate message, the PCC should send the prefixes indicated in this object to the appointed BGP peer.

When PCC sends successfully the prefixes to the appointed BGP peer, it should report the result via the PCRpt messages, with PPA object included, and the corresponding SRP and CCI object.

When PCC receives the PPA and the CCI object with the R bit set to 1 in SRP object in PCInitiate message, the PCC should withdraw the prefixes advertisement to the peer that indicated by this object.

When PCC withdraws successfully the prefixes that indicated by this object, it should report the result via the PCRpt message, with the PPA object included, and the corresponding SRP and CCI object.

The IPv4 prefix MUST only be advertised via the IPv4 BGP session and the IPv6 prefix MUST only be advertised via the IPv6 BGP session. If mismatch occur, an error(Error-type=TBD6, Error-value=TBD18, BPI/PPR address family mismatch) should be reported via PCErr message.

When the peer info that associated with the Symbolic Path Name is not the same as the peer info that indicated in BPI object in PCC, an error (Error-type=TBD6, Error-value=TBD19, PPA/BPI peer info mismatch) should be reported via the PCErr message.

```
                 +------------------+
       +----------+       PCE        +-----------+
       |          +------------------+           |
       |                  +--+                   |
       +------------------+R3+-------------------+
     M1&M1-R              +--+                M2&M2-R
       |                                         |
       +v-+          +--+          +--+         +-v+
       |R1+----------+R5+----------+R6+---------+R7|
       ++-+          +--+          +--+         +-++
       |                                         |
       |                                         |
       |            +--+          +--+           |
       +------------+R2+----------+R4+-----------+
```

Figure 4: BGP Prefix Advertisement Procedures

Table 4: Message Information

No.	Peers	Type	Message Key Parameters
M1	PCE/R1	PCInitiate	CC-ID=X1(Symbolic Path Name=Class A)
M1-R		PCRpt	PPA Object(Peer IP=R7_A,Prefix=1_A)
M2	PCE/R7	PCInitiate	CC-ID=X2(Symbolic Path Name=Class A)
M2-R		PCRpt	PPA Object(Peer IP=R1_A,Prefix=7_A)

7. New PCEP Objects

One new CCI Object and three new PCEP objects are defined in this draft. All new PCEP objects are as per [RFC5440]

7.1. 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]. This document defines another object-type for Native-IP.

CCI Object-Type is TBD13 for Native-IP 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                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Reserved             |            Flags              |
+---------------------------------------------------------------+
|                                                               |
//                        Optional TLV                         //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: CCI Object for Native IP

Figure 1

The field CC-ID is as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. Following fields are defined for CCI Object-Type TBD13

Reserved: is set to zero while sending, ignored on receipt.

Flags: is used to carry any additional information pertaining to the CCI. Currently no flag bits are defined.

The Symbolic Path Name TLV [RFC8231] MUST be included in the CCI Object-Type TBD13 to identify the E2E TE path in Native IP environment and MUST be unique.

7.2. BGP Peer Info Object

The BGP Peer Info object is used to specify the information about the peer that the PCC should establish the BGP relationship with. This object should only be included and sent to the head and end router of the E2E path in case there is no Route Reflection (RR) involved. If the RR is used between the head and end routers, then such information should be sent to head router, RR and end router respectively.

By default, there MUST be no prefix be distributed via such BGP session that established by this object.

By default, the Local/Peer IP address SHOULD be dedicated to the usage of native IP TE solution, and SHOULD NOT be used by other BGP sessions that established by manual or non PCE initiated configuration.

BGP Peer Info Object-Class is TBD14

BGP Peer Info Object-Type is 1 for IPv4 and 2 for IPv6

The format of the BGP Peer Info object body for IPv4(Object-Type=1) 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Peer AS Number                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   ETTL        |             Reserved                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Local IP Address                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Peer IP Address                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Additional TLVs                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: BGP Peer Info Object Body Format for IPv4

The format of the BGP Peer Info object body for IPv6(Object-Type=2) 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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                      Peer AS Number                           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   ETTL        |  Tunnel Type  |        Reserved               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                                                               +
    |               Local IP Address (16 bytes)                     |
    +                                                               +
    |                                                               |
    +                                                               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                                                               +
    |               Peer IP Address (16 bytes)                      |
    +                                                               +
    |                                                               |
    +                                                               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                  Additional TLVs                              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 7: BGP Peer Info Object Body Format for IPv6

Peer AS Number: 4 Bytes, to indicate the AS number of Remote Peer.

ETTL: 1 Byte, to indicate the multi hop count for EBGP session. It should be 0 and ignored when Local AS and Peer AS is same.

Tunnel Type: 1 Byte, indicate the tunnel type that used to transfer the traffic that identified by the prefixes that advertised by the corresponding BGP peer. Value 0 indicate no tunnel is used. Other value can refer to the IANA allocation value in "BGP Tunnel Encapsulation Attribute Tunnel Types".

Reserved: is set to zero while sending, ignored on receipt..

Local IP Address(4/16 Bytes): IP address of the local router, used to peer with other end router. When Object-Type is 1, length is 4 bytes; when Object-Type is 2, length is 16 bytes.

Peer IP Address(4/16 Bytes): IP address of the peer router, used to peer with the local router. When Object-Type is 1, length is 4 bytes; when Object-Type is 2, length is 16 bytes;

Additional TLVs: TLVs that associated with this object, can be used to convey other necessary information for dynamic BGP session establishment. Its definition is out of the current document.

When PCC receives BPI object, with Object-Type=1, it should try to establish BGP session with the peer in AFI/SAFI=1/1; when PCC receives BPI object with Object-Type=2, it should try to establish the BGP session with the peer in AFI/SAFI=2/1. Other BGP capabilities,for example, Graceful Restart(GR) that enhance the BGP performance should also be negotiated and used by default.

7.3. Explicit Peer Route Object

The Explicit Peer Route object is defined to specify the explicit peer route to the corresponding peer address on each device that is on the E2E assurance path. This Object should be sent to all the devices that locates on the E2E assurance path that calculated by PCE.

The path established by this object should have higher priority than other path calculated by dynamic IGP protocol, but should be lower priority that the static route configured by manual or NETCONF channel.

Explicit Peer Route Object-Class is TBD15.

Explicit Peer Route Object-Type is 1 for IPv4 and 2 for IPv6

The format of Explicit Peer Route object body for IPv4(Object-Type=1) 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Route Priority          |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 IPv4 Peer Address                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Next Hop Address to the IPv4 Peer Address          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Additional TLVs                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8: Explicit Peer Route Object Body Format for IPv4

The format of Explicit Peer Route object body for IPv6(Object-Type=2) 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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |       Route Priority          |           Reserved            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                                                               +
    |                       IPv6 Peer Address                       |
    +                                                               +
    |                                                               |
    +                                                               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                                                               +
    |          Next Hop Address to the IPv6 Peer Address            |
    +                                                               +
    |                                                               |
    +                                                               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                  Additional TLVs                              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 9: Explicit Peer Route Object Body Format for IPv6

Route Priority: 2 Bytes, The priority of this explicit route. The higher priority should be preferred by the device. This field is used to indicate the backup path at each hop.

Reserved.: is set to zero while sending, ignored on receipt.

Peer Address: To indicate the peer address.

Next Hop Address to the Peer: To indicate the next hop address to the corresponding peer.

Additional TLVs: TLVs that associated with this object, can be used to convey other necessary information for explicit peer path establishment. Its definition is out of the current document.

7.4. Peer Prefix Association Object

The Peer Prefix Association object is defined to specify the IP prefixes that should be advertised to the corresponding peer. This object should only be included and sent to the head/end router of the end2end path.

The prefixes information included in this object MUST only be advertised to the indicated peer, MUST NOT be advertised to other BGP peers.

Peer Prefix Association Object-Class is TBD16

Peer Prefix Association Object-Type is 1 for IPv4 and 2 for IPv6

The format of the Peer Prefix Association 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Peer IPv4 Address                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//               IPv4 Prefix subobjects                        //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Additional TLVs                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 10: Peer Prefix Association Object Body Format for IPv4

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Peer IPv6 Address                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//               IPv6 Prefix subobjects                        //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Additional TLVs                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 11: Peer Prefix Association Object Body Format for IPv6

Peer IPv4 Address: 4 Bytes. Identifies the peer IPv4 address that the associated prefixes will be sent to.

IPv4 Prefix subojects: List of IPv4 Prefix subobjects that defined in [RFC3209], identify the prefixes that will be sent to the peer that identified by Peer IPv4 Address List.

Peer IPv6 Address: 16 Bytes. Identifies the peer IPv6 address that the associated prefixes will be sent to.

IPv6 Prefix subojects: List of IPv6 Prefix subobjects that defined in [RFC3209], identify the prefixes that will be sent to the peer that identified by Peer IPv6 Address List.

Additional TLVs: TLVs that associated with this object, can be used to convey other necessary information for prefixes advertisement. Its definition is out of the current document.

8. End to End Path Protection

[RFC8697] defines the path associations procedures between sets of Label Switched Path (LSP). Such procedures can also be used for the E2E path protection. To accomplish this, the PCE should attach the ASSOCIATION object with the EPR object in the PCInitiate message, with the association type set to 1 (Path Protection Association). The Extended Association ID that included within the Extended Association ID TLV, which is included in the ASSOCIATION object, should be set to the Symbolic Path Name of different E2E path. This PCinitiate should be sent to the head-end of the E2E path.

The head-end of the path can use the existing path detection mechanism, to monitor the status of the active path. Once it detects the failure, it can switch the backup protection path immediately.

9. New Error-Types and Error-Values Defined

A PCEP-ERROR object is used to report a PCEP error and is characterized by an Error-Type that specifies that type of error and an Error-value that provides additional information about the error. An additional Error-Type and several Error-values are defined to represent some the errors related to the newly defined objects, which are related to Native IP TE procedures.

Error-Type	Meaning	Error-value
TBD6	Native IP TE failure	
		0: Unassigned
		TBD7: Peer AS not match
		TBD8:Peer IP can't be reached
		TBD9:Local IP is in use
		TBD10:Remote IP is in use
		TBD11:Exist BGP session broken
		TBD12:Explicit Peer Route Error
		TBD17:EPR/BPI Peer Info mismatch
		TBD18:BPI/PPA Address Family mismatch
		TBD19:PPA/BPI Peer Info mismatch

Figure 12: Newly defined Error-Type and Error-Value

10. Deployment Considerations

The information transferred in this draft is mainly used for the light weight BGP session setup, explicit route deployment and the prefix distribution. The planning, allocation and distribution of the peer addresses within IGP should be accomplished in advanced and they are out of the scope of this draft.

[RFC8232] describes the state synchronization procedure between stateful PCE and PCC. The communication of PCE and PCC described in this draft should also follow this procedures, treat the three newly

defined objects that associated with the same symbolic path name as the attribute of the same path in the LSP-DB.

When PCE detects one or some of the PCCs are out of control, it should recompute and redeploy the traffic engineering path for native IP on the active PCCs. When PCC detects that it is out of control of the PCE, it should clear the information that initiated by the PCE. The PCE should assures the avoidance of possible transient loop in such node failure when it deploy the explicit peer route on the PCCs.

If the established BGP session is broken after some time, the PCC should also report such error via PCErr message with Err-type=TBD6 and error value(Error-value=TBD11, Existing BGP session is broken). Upon receiving such PCErr message, the PCE should clear the prefixes advertisement on the previous BGP session, clear the explicit peer route to the previous peer address; select other Local_IP/Peer_IP pair to establish the new BGP session, deploy the explicit peer route to the new peer address, and advertises the prefixes on the new BGP session.

11. Security Considerations

Service provider should consider the protection of PCE and their communication with the underlay devices, which is described in document [RFC5440] and [RFC8253]

12. IANA Considerations

12.1. Path Setup Type Registry

[RFC8408] created a sub-registry within the "Path Computation Element Protocol (PCEP) Numbers" registry called "PCEP Path Setup Types". IANA is requested to allocate a new code point within this registry, as follows:

Value	Description	Reference
TBD1	Native IP TE Path	This document

12.2. PCECC-CAPABILITY sub-TLV's Flag field

[I-D.ietf-pce-pcep-extension-for-pce-controller] created a sub-registry within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the value of the PCECC-CAPABILITY sub-TLV's 32-bits Flag field. IANA is requested to allocate a new bit position within this registry, as follows:

Value	Description	Reference
TBD2(N)	NATIVE-IP-TE-CAPABILITY	This document

12.3. PCEP Object Types

IANA is requested to allocate new registry for the PCEP Object Type:

```
Object-Class Value       Name                            Reference
TBD13             CCI Object                        This document
                  Object-Type
                     TBD: Native IP

TBD14             BGP Peer Info                     This document
                  Object-Type
                     1: IPv4 address
                     2: IPv6 address

TBD15             Explicit Peer Route               This document
                  Object-Type
                     1: IPv4 address
                     2: IPv6 address

TBD16             Peer Prefix Association           This document
                  Object-Type
                     1: IPv4 address
                     2: IPv6 address
```

12.4. 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 value	Meaning	Error-value	Reference
6	**Mandatory Object missing**	**TBD4:Native IP object missing**	
10	**Reception of an invalid object**	**TBD3:PCECC NATIVE-IP-TE-CAPABILITY bit is not set**	
19	**Invalid Operation**	**TBD5:Only one of the BPI,EPR or PPA object can be included**	
TBD6	Native IP TE failure		This document
		TBD7:Peer AS not match	
		TBD8:Peer IP can't be reached	
		TBD9:Local IP is in use	
		TBD10:Remote IP is in use	
		TBD11:Exist BGP session broken	
		TBD12:Explicit Peer Route Error	
		TBD17:EPR/BPI Peer Info mismatch	
		TBD18:BPI/PPA Address Family mismatch	
		TBD19:PPA/BPI Peer Info mismatch	

13. Contributor

Dhruv Dhody has contributed the contents of this draft.

14. Acknowledgement

Thanks Mike Koldychev, Siva Sivabalan, Adam Simpson for his valuable suggestions and comments.

15. Normative References

[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-10 (work in progress), January 2021.

[I-D.ietf-teas-pce-native-ip]
Wang, A., Khasanov, B., Zhao, Q., and H. Chen, "Path Computation Element (PCE) based Traffic Engineering (TE) in Native IP Networks", draft-ietf-teas-pce-native-ip-16 (work in progress), January 2021.

[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>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[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>.

[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>.

[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>.

[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>.

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

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

[RFC8735] Wang, A., Huang, X., Kou, C., Li, Z., and P. Mi, "Scenarios and Simulation Results of PCE in a Native IP Network", RFC 8735, DOI 10.17487/RFC8735, February 2020, <https://www.rfc-editor.org/info/rfc8735>.

Authors' Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing, Beijing 102209
China

Email: wangaj3@chinatelecom.cn

Boris Khasanov
Yandex LLC
Ulitsa Lva Tolstogo 16
Moscow
Russia

Email: bhassanov@yahoo.com

Sheng Fang
Huawei Technologies,Co.,Ltd
Huawei Bld., No.156 Beiqing Rd.
Beijing
China

Email: fsheng@huawei.com

Ren Tan
Huawei Technologies,Co.,Ltd
Huawei Bld., No.156 Beiqing Rd.
Beijing
China

Email: tanren@huawei.com

Chun Zhu
ZTE Corporation
50 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: zhu.chun1@zte.com.cn