PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: 5 August 2024

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

2 February 2024

Path Computation Element Communication Protocol (PCEP) Extensions for Native IP Networks

draft-ietf-pce-pcep-extension-native-ip-30

Abstract

This document defines the Path Computation Element Communication Protocol (PCEP) extension for Central Control Dynamic Routing (CCDR) based applications in Native IP networks. It 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 the Native IP network under PCE as a central controller (PCECC).

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Generally, Multiprotocol Label Switching Traffic Engineering (MPLS-TE) requires the corresponding network devices to support Resource ReSerVation Protocol (RSVP)/Label Distribution Protocol (LDP) protocols to assure the End-to-End (E2E) traffic performance. But in native IP network scenarios described in [RFC8735], there will be no such signaling protocol to synchronize the actions among different network devices. It is feasible to use the central control mode described in [RFC8283] to correlate the forwarding behavior among different network devices. [RFC8821] describes the architecture and solution philosophy for the E2E traffic assurance in the Native IP network via multiple Border Gateway Protocol (BGP) session-based solution. It requires only the PCE send the instructions to the PCCs, to build multiple BGP sessions, distribute different prefixes on the established BGP sessions and assign the different paths to the BGP next hops.

This document describes the corresponding Path Computation Element Communication Protocol (PCEP) extensions to transfer the key information about BGP peer, peer prefix advertisement, 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]: PCC, PCE, PCEP.

The following terminology is used in this document:

* CCDR: Central Control Dynamic Routing

* E2E: End-to-End

* BPI: BGP Peer Info

* EPR: Explicit Peer Route

* PPA: Peer Prefix Advertisement

* PST: Path Setup Type

* PCECC: PCE as a Central Controller

* RR: Route Reflector

4. Capability Advertisement

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:

* PST = 4: Path is a Native IP TE path as per [RFC8821].

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.

[RFC9050] 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 - 30): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable of TE in a Native IP network as specified in this document. The flag MUST be set by both the PCC and PCE 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:

- send a PCErr message with Error-Type=10 (Reception of an invalid object) and Error-Value=39 (PCECC NATIVE-IP-TE-CAPABILITY bit is not set).

- terminate the PCEP session

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

- send a PCErr message with Error-Type=10(Reception of an invalid object) and Error-Value=33 (Missing PCECC Capability sub-TLV).

- terminate the PCEP session

If one or both speakers (PCE and PCC) have not indicated support and willingness to use the PCEP extensions for Native-IP, the PCEP extensions for the Native-IP MUST NOT be used. If a Native-IP operation is attempted when both speakers have not agreed in the OPEN messages, the receiver of the message MUST:

- send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD1 (Attempted Native-IP operations when capability was not advertised) and

- terminate the PCEP session.

5. PCEP Messages

PCECC Native IP TE solution uses the existing PCE Label Switched Path (LSP) Initiate Request message (PCInitiate) [RFC8281], and PCE Report message (PCRpt) [RFC8231] to accomplish the multiple BGP sessions establishment, E2E Native-IP 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 [RFC9050] is used to download or remove central controller’s instructions (CCIs). [RFC9050] specifies an object called CCI for the encoding of the central controller’s instructions. This document specifies 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 Advertisement (PPA) Object) are defined in this document. Refer to Section 7 for detailed object definitions. All PCEP procedures specified in [RFC9050] continue to apply unless specified otherwise.

5.1. The PCInitiate Message

The PCInitiate Message defined in [RFC8281] and extended in [RFC9050] 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>

     <cci-list> ::=  <CCI>
                     [<BPI>|<EPR>|<PPA>]
                     [<cci-list>]
```

Where:

<PCE-initiated-lsp-instantiation> and <PCE-initiated-lsp-deletion> are as per [RFC8281].

The LSP and SRP objects are defined in [RFC8231].

When PCInitiate message is used for Native IP instructions, the SRP, LSP and CCI objects MUST be present. The error handling for missing SRP, LSP or CCI object is as per [RFC9050]. Further only one object among BPI, EPR, or PPA object MUST be present. The PLSP-ID and Symbolic Path Name TLV are set as per the existing rules in [RFC8231], [RFC8281], and [RFC9050]. The Symbolic Path Name is used by the PCE/PCC to uniquely identify the E2E native IP TE path. The related Native-IP instructions with BPI, EPR and PPA object are identified by the same Symbolic Path Name.

If none of the BPI, EPR, or PPA object are present, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=19 (Native IP object missing). If there are more than one instance of BPI, EPR or PPA object present, the receiving PCC MUST send a PCErr message with Error-type=19 (Invalid Operation) and Error-value=22 (Only one BPI, EPR or PPA object can be included in this message).

To cleanup the existing Native IP instructions, 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) as well as during the State Synchronization phase.

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>

      <cci-list> ::=  <CCI>
                     [<BPI>|<EPR>|<PPA>]
                     [<cci-list>]

   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 [RFC9050]. Further only one object among BPI, EPR, or PPA object MUST be present.

If none of the BPI, EPR, or PPA object are present, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=19 (Native IP object missing). If there are more than one instance of BPI, EPR or PPA object present, the receiving PCE MUST send a PCErr message with Error-type=19 (Invalid Operation) and Error-value=22 (Only one 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 PCInitiate and PCRpt message pair is used to exchange the configuration parameters for a BGP peer session. This pair of PCEP messages is exchanged between a PCE and each BGP peer (acting as PCC) which needs to establish BGP session. After the BGP peer session has been initiated via this pair of PCEP messages, the BGP session establishes and operates in a normal fashion. The BGP peers can be used for External BGP (EBGP) peers or Internal BGP (IBGP) peers. For IBGP connection topologies, the Route Reflector (RR) is required.

The PCInitiate message should be sent to PCC which is acting as BGP router and/or RR.

The RR topology for a single Autonomous System (AS) is shown in Figure 1. The BGP routers R1, R3, and R7 are within a single AS. R1 and R7 are BGP RR clients, and R3 is a RR. The PCInitiate message should be sent to the BGP routers R1, R3 and R7 that need to establish BGP session .

PCInitiate message creates an auto-configuration function for these BGP peers by providing the indicated Peer AS and the Local/Peer IP Address.

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 per AS and Local/Peer IP address.

During the establishment procedure, PCC should report to the PCE the status of the BGP session via the PCRpt message, with the status field in the BPI object set to the appropriate value and the corresponding SRP and CCI object included.

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 is 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 objects.

```
                                +--------------------+
                    +----------->       PCE        <----------+
                    |           +--------^---------+          |
                    |                    |                    |
                    |             PCInitiate/PCRpt            |
                    |                    |                    |
                    |               +----v--+                 |
                    +---------------+ R3(RR)+-----------------+
                    |               +-------+                 |
              PCInitiate/PCRpt                         PCInitiate/PCRpt
                    |                                         |
                   +v-+          +--+          +--+         +-v+
                   |R1+----------+R5+----------+R6+---------+R7|
                   ++-+          +-++          +--+         +-++
                    |              |                          |
                    |            +--+          +--+           |
                    +------------+R2+----------+R4+-----------+
                                 +--+          +--+
```

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

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

```
               +-------+                                       +--------+
               |PCC    |                                       |  PCE   |
               |R1     |                                       +--------+
        +------|       |                                            |
        | PCC  +-------+                                            |
        | R3     | |   (For R1/R3 BGP Session on R1)                |
 +------|        | |<-PCInitiate,CC-ID=X,Symbolic Path Name=Class A-|
 |      |        | |BPI Object(Peer AS, Local_IP=R1_A, Peer_IP=R3_A)|
 |PCC   +--------+ |                                                |
 |R7      |  |     |----PCRpt,CC-ID=X(Symbolic Path Name=Class A)-->|
 |        |  |     |BPI Object(Peer AS, Local_IP=R1_A, Peer_IP=R3_A)|
 +--------+  |                                                      |
     |       |          (For R1/R3 BGP Session on R3)               |
     |       |<--PCInitiate,CC-ID=Y1,Symbolic Path Name=Class A-----|
     |       |      BPI Object(Peer AS, Local_IP=R3_A, Peer_IP=R1_A)|
     |       |---PCRpt,CC-ID=Y1,Symbolic Path Name=Class A--------->|
     |       |      BPI Object(Peer AS, Local_IP=R3_A, Peer_IP=R1_A)|
     |       |                                                      |
     |       |          (For R3/R7 BGP Session on R3)               |
     |       |<--PCInitiate,CC-ID=Y2,Symbolic Path Name=Class A-----|
     |       |  BPI Object(Peer AS, Local_IP=R3_A, Peer_IP=R7_A)    |
     |       |----PCRpt,CC-ID=Y2,Symbolic Path Name=Class A-------->|
     |       |  BPI Object(Peer AS, Local_IP=R3_A, Peer_IP=R7_A)    |
     |                                                              |
     |                  (For R3/R7 BGP Session on R7)               |
     |<--PCInitiate,CC-ID=Z,Symbolic Path Name=Class A--------------|
     |            BPI Object(Peer AS, Local_IP=R7_A, Peer_IP=R3_A)  |
     |---PCRpt,CC-ID=Z,Symbolic Path Name=Class A------------------>|
     |            BPI Object(Peer AS, Local_IP=R7_A, Peer_IP=R3_A)  |
```

Figure 2: Message Information and Procedures

The Local/Peer IP address MUST be dedicated to the usage of native IP TE solution, and MUST NOT be used by other BGP sessions that established by manual or other ways. If the Local IP Address or Peer IP Address within BPI object is used in other existing BGP sessions, the PCC SHOULD report such error situation via a PCErr message with:

Error-type=33 (Native IP TE failure) and Error-value=1 (Local IP is in use), or

Error-type=33 (Native IP TE failure)and Error-value=2 (Remote IP is in use).

The detailed Error-Types and Error-Values are defined in Section 8

If the established BGP session is broken, the PCC should report such information via PCRpt message with the status field set to "BGP session down" in associated BPI Object. The error code field within the BPI object should indicate the reason that leads to the BGP session down. In future, when the BGP session is up again, the PCC should report that as well via the PCRpt message with status field set to "BGP Session Established".

6.2. Explicit Route Establishment Procedures

The explicit route establishment procedures can be used to install a route via PCE on the PCC, using PCInitiate and PCRpt message pair. Such explicit routes operate the same as static routes installed by network management protocols (Network Configuration Protocol (NETCONF)/YANG). The procedures of such explicit route addition and remove must be controlled by the PCE in an specific order so that the pathways are established without loops.

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, R2 and R4, as shown in Figure 3. For explicit route from R7 to R1, the PCInitiate message should be sent to R7, R4 and R2, as shown in Figure 5.

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 in the RIB/FIB to the peer.

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

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 is indicated by the EPR object.

When PCC has cleared the explicit route that is 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)+--|-------------+
             PCInitiate/PCRpt     | +------+  |             |
                  |               |           |             |
                 +v-+      +--+   |           |   +--+    +--+
                 |R1+------+R5+---+-----------|---+R6+----+R7|
                 ++-+      +--+   |           |   +--+    +-++
                  |     PCInitiate/PCRpt  PCInitiate/PCRpt  |
                  |               |           |             |
                  |            +--v--+     +--v-+           |
                  +------------+- R2 +-----+ R4 +-----------+
                               +--+--+     +--+-+
```

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

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

```
               +-------+                                       +-------+
               |PCC    |                                       |  PCE  |
               |R4     |                                       +-------+
        +------|       |                                           |
        | PCC  +-------+                                           |
        | R2     | |        (EPR route on R4)                      |
 +------|        | |<-PCInitiate,CC-ID=Z,Symbolic Path Name=Class A|
 |      |        | |   EPR Object(Peer Address=R7_A, Next Hop=R7_A)|
 |PCC   +--------+ |                                               |
 |R1      |  |     |----PCRpt,CC-ID=Z,Symbolic Path Name=Class A-->|
 |        |  |     |   EPR Object(Peer Address=R7_A, Next Hop=R7_A)|
 +--------+  |                                                     |
     |       |              (EPR route on R2)                      |
     |       |<--PCInitiate,CC-ID=Y,Symbolic Path Name=Class A-----|
     |       |   EPR Object(Peer Address=R7_A, Next Hop=R4_A)      |
     |       |----PCRpt,CC-ID=Y,Symbolic Path Name=Class A-------->|
     |       |   EPR Object(Peer Address=R7_A, Next Hop=R4_A)      |
     |       |                                                     |
     |                                                             |
     |                      (EPR route on R1)                      |
     |<--PCInitiate,CC-ID=X,Symbolic Path Name=Class A-------------|
     |              EPR Object(Peer Address=R7_A, Next Hop=R2_A)   |
     |---PCRpt,CC-ID=X1(Symbolic Path Name=Class A)---------------->|
     |              EPR Object(Peer Address=R7_A, Next Hop=R2_A)   |
```

Figure 4: Message Information and Procedures

```
                       +------------------+
                       +       PCE        <-----------+
                       +----^-----------^-+           |
                            |           |             |
                            |           |             |
                            |           |             |
                            | +------+  |             |
            +-----------------+R3(RR)+--|-------------+
            |               | +------+  |       PCInitiate/PCRpt
            |               |           |             |
           +--+      +--+   |           |   +--+    +-v+
           |R1+------+R5+---+-----------|---+R6+----+R7|
           ++-+      +--+   |           |   +--+    +-++
            |       PCInitiate/PCRpt PCInitiate/PCRpt |
            |               |           |             |
            |            +--v--+     +--v-+           |
            +------------+- R2 +-----+ R4 +-----------+
                         +--+--+     +--+-+
```

Figure 5: Explicit Route Establish Procedures(From R7 to R1)

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

```
               +-------+                                       +-------+
               |PCC    |                                       |  PCE  |
               |R2     |                                       +-------+
        +------|       |                                           |
        | PCC  +-------+                                           |
        | R4       | |        (EPR route on R2)                      |
 +------|        | |<-PCInitiate,CC-ID=X,Symbolic Path Name=Class A|
 |      |        | |  EPR Object(Peer Address=R1_A, Next Hop=R1_A) |
 |PCC   +--------+ |                                               |
 |R7      |  |     |----PCRpt,CC-ID=X,Symbolic Path Name=Class A-->|
 |        |  |     |  EPR Object(Peer Address=R1_A, Next Hop=R1_A) |
 +--------+  |                                                     |
     |       |              (EPR route on R4)                      |
     |       |<--PCInitiate,CC-ID=Y,Symbolic Path Name=Class A-----|
     |       |   EPR Object(Peer Address=R1_A, Next Hop=R2_A)      |
     |       |----PCRpt,CC-ID=Y,Symbolic Path Name=Class A-------->|
     |       |   EPR Object(Peer Address=R1_A, Next Hop=R2_A)      |
     |       |                                                     |
     |                                                             |
     |                      (EPR route on R7)                      |
     |<--PCInitiate,CC-ID=Z,Symbolic Path Name=Class A-------------|
     |   EPR Object(Peer Address=R1_A, Next Hop=R4_A)              |
     |---PCRpt,CC-ID=Z,Symbolic Path Name=Class A----------------->|
     |   EPR Object(Peer Address=R1_A, Next Hop=R4_A)              |
```

Figure 6: Explicit Route Establish Procedures(From R7 to R1)

In order to avoid the transient loop while deploying the 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 the E2E path.

To accomplish ECMP effects, the PCE can send multiple EPR/CCI objects to the same node, with the same route priority and peer address value but different next hop address.

The PCC should verify that the next hop address is reachable. In case of failure, the PCC SHOULD send the corresponding error via PCErr message, with an error information: Error-type=33 (Native IP TE failure), Error-value=3 (Explicit Peer Route Error).

When the peer info is not the same as the peer info that is indicated in the BPI object in PCC for the same path that is identified by Symbolic Path Name TLV, an PCErr message SHOULD be reported, with an error information: Error-type=33 (Native IP TE failure), Error-value=4, EPR/BPI Peer Info Mismatch. Note that the same error can be used in case no BPI is received at the PCC.

If the PCE needs to update the path, it should first instruct new CCI with updated EPR corresponding to the new next hop to use and then instruct the removal of older CCI.

6.3. BGP Prefix Advertisement Procedures

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

The PCInitiate message should be sent to PCC that acts as BGP peer edge router only. In the example, it should be sent to R1 and R7 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 identified BGP peer via the corresponding BGP session [RFC4271].

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

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 indicated by this object.

When PCC withdraws successfully the prefixes that is 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.

```
                    +------------------+
         +---------->       PCE        <-----------+
         |          +------------------+           |
         |                  +--+                   |
         +------------------+R3+-------------------+
   PCInitiate/PCRpt         +--+             PCInitiate/PCRpt
         |                                         |
        +v-+          +--+          +--+         +-v+
        |R1+----------+R5+----------+R6+---------+R7|
        ++-+          +--+          +--+         +-++
    (BGP Router)                               (BGP Router)
         |                                         |
         |                                         |
         |            +--+          +--+           |
         +------------+R2+----------+R4+-----------+
                      +--+          +--+
```

Figure 7: BGP Prefix Advertisement Procedures

```
                +-------+                                      +-------+
                |PCC    |                                      |  PCE  |
                |R1     |                                      +-------+
         +------|       |                                           |
         | PCC  +-------+                                           |
         | R7     | |   (Instruct R1 to advertise Prefix 1_A to R7) |
         |        | |<-PCInitiate,CC-ID=X,Symbolic Path Name=Class A|
         |        | |  PPA Object(Peer IP=R7_A, Prefix=1_A)         |
         +--------+ |                                               |
              |     |                                               |
              |     |----PCRpt,CC-ID=X,Symbolic Path Name=Class A-->|
              |     |    PPA Object(Peer IP=R7_A, Prefix=1_A)       |
              |                                                     |
              |     (Instruct R7 to advertise Prefix 7_A to R1 )    |
              |<--PCInitiate,CC-ID=Z,Symbolic Path Name=Class A-----|
              |         PPA Object(Peer IP=R1_A, Prefix=7_A)        |
              |----PCRpt,CC-ID=Z,Symbolic Path Name=Class A-------->|
              |              PPA Object(Peer IP=R1_A, Prefix=7_A)   |
              |                                                     |
```

Figure 8: Message Information and Procedures

The AFI/SAFI for the corresponding BGP session should match the Peer Prefix Advertisement Object-Type, AFI/SAFI should be 1/1 for IPv4 prefix and 2/1 for IPv6 prefix. In case of mismatch, an error: Error-type=33 (Native IP TE failure), Error-value=5 (BPI/PPA address family mismatch) SHOULD be reported via PCErr message.

When the peer info is not the same as the peer info that is indicated in the BPI object in PCC for the same path that is identified by Symbolic Path Name TLV, an error: Error-type=33 (Native IP TE failure), Error-value=6 (PPA/BPI peer info mismatch) SHOULD be reported via the PCErr message. Note that the same error can be used in case no BPI is received at the PCC.

6.4. Selection of Raw Mode and Tunnel Mode forwarding strategy

Normally, when the above procedures are finished, the user traffic will be forwarded via the appointed path, but the forwarding will be based solely on the destination of user traffic. If there are traffic from different attached point to the same destination coming into the network, they could share the priority path which may not be the initial desire. For example, as illustrated in Figure 1, the initial aim is to assure traffic that enter into the network via R1, and exit the network at R7 via R5-R6-R7. If some traffic enter into the network via the R2 router, pass through R5 and exit at R7, they may share the priority path among R5-R6-R7, which may not be the desired effect.

The above normal traffic forwarding behaviour are clarified as Raw mode forwarding strategy. Such mode can achieve only the moderate traffic path control effect. In order to achieve the strict traffic path control effect, the entry point should tunnel the user traffic from the entry point of the network to the exit point of the network, which is also between the BGP peer that established via Section 6.1. Such forwarding behavior are called Tunnel mode forwarding strategy.

The selection of Raw mode and Tunnel mode forwarding strategy are controlled via the "T" bit in BPI Object that is defined in Section 7.2

6.5. Cleanup

In order to remove the Native-IP state from the PCC, the PCE MUST send explicit CCI cleanup instruction for PPA, EPR, and BPI object respectively with R flag set in the SRP object. If the PCC receives a PCInitiate message but does not recognize the Native-IP information in the CCI, the PCC MUST generate a PCErr message with Error-Type=19 (Invalid operation) and Error-value=TBD2 (Unknown Native-IP Info) and MUST include the SRP object to specify the error is for the corresponding cleanup (via a PCInitiate message).

6.6. Other Procedures

The handling of the state synchronization, redundant PCEs, re-delegation and clean up is the same as other CCIs as specified in [RFC9050].

7. New PCEP Objects

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

7.1. CCI Object

The Central Control Instructions (CCI) Object (defined in [RFC9050]) is used by the PCE to specify the forwarding instructions. This document defines another object-type for Native-IP procedures.

CCI Object-Type is 2 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 9: CCI Object for Native IP

The field CC-ID is as described in [RFC9050]. Following fields are defined for CCI Object-Type 2

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

Flags: is used to carry any additional information pertaining to the Native-IP CCI. Currently no flag bits are defined. Unassigned flags are set to zero while sending and ignored on receipt.

The Symbolic Path Name TLV [RFC8231] MUST be included in the CCI Object-Type 2 to identify the E2E TE path in Native IP environment.

7.2. BGP Peer Info Object

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

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 other configuration mechanism.

BGP Peer Info Object-Class is 46

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        |     Status    |   Error Code  |    Flag     |T|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Local IP Address                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Peer IP Address                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //                    Optional TLVs                            //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 10: 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        |      Status   |  Error Code   |    Flag     |T|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |               Local IP Address (16 bytes)                     |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |               Peer IP Address (16 bytes)                      |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //                    Optional TLVs                            //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 11: BGP Peer Info Object Body Format for IPv6

Peer AS Number: 4 Bytes, to indicate the AS number of Remote Peer. Note that if 2-byte AS numbers are in use, the low-order bits (16 through 31) MUST be used, and the high-order bits (0 through 15) MUST be set to zero.

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

Status: 1 Byte, Indicate BGP session status between the peers. It’s values are defined below:

- 0: Reserved

- 1: BGP Session Established

- 2: BGP Session Establishment In Progress

- 3: BGP Session Down

- 4-255: Reserved

Error Code: 1 Byte, Indicate the reason that BGP session can’t be established.

- 0: Reserved

- 1: ASes does not match, BGP Session Failure

- 2: Peer IP can’t be reached, BGP Session Failure

- 3-255: Reserved

Flag: 1 Byte.

- Currently only bit 7 (T bit) is defined. When T bit is set, the traffic should be sent in IPinIP tunnel (Tunnel source is Local IP Address, tunnel destination is Peer IP Address). When T bit is cleared, the traffic is sent via its original source and destination address. The Tunnel mode(T bit is set) is used when the operator want to assure only the traffic from the specified (entry, exit) pair, the Raw mode (T bit is clear) is used when the operator want to assure traffic from any entry to the specified destination. Unassigned flags are set to zero while sending and 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;

Optional TLVs: TLVs that associated with this object, can be used to convey other necessary information for dynamic BGP session establishment. No TLVs are currently defined.

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.

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 Native-IP TE path. This Object should be sent to all the devices on the path that is calculated by the PCE.

It is RECOMMENDED that the path established by this object should have higher priority than the other paths calculated by dynamic IGP protocol, but should have lower priority than the static route configured by manual or NETCONF or any other static means.

Explicit Peer Route Object-Class is 47.

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              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Peer IPv4 Address                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |               Next Hop IPv4 Address to the Peer                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //                    Optional TLVs                             //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 12: 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             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |                       Peer IPv6 Address                        |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |                Next Hop IPv6 Address to the Peer               |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //                    Optional TLVs                             //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 13: 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 preferred path at each hop.

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

Peer (IPv4/IPv6) Address: Peer Address for the BGP session (4/16 Bytes).

Next Hop (IPv4/IPv6) Address to the Peer: To indicate the next hop address (4/16 Bytes) to the corresponding peer address.

Optional TLVs: TLVs that associated with this object, can be used to convey other necessary information for explicit peer path establishment. No TLVs are currently defined.

7.4. Peer Prefix Advertisement Object

The Peer Prefix Advertisement 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 source/destination router of the E2E 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 Advertisement Object-Class is 48

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

The format of the Peer Prefix Advertisement 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                            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | No. of Prefix |                  Reserved                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                  IPv4 Prefix #1                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Prefix #1 Len  |                  Reserved                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                               :                               |
    |                               :                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                  IPv4 Prefix #n                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Prefix #n Len  |                  Reserved                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //                     Optional TLVs                           //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 14: Peer Prefix Advertisement 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                            |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | No. of Prefix |                  Reserved                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  IPv6 Prefix #1                               |
   |                                                               |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Prefix #1 Len  |                  Reserved                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               :                               |
   |                               :                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  IPv6 Prefix #n                               |
   |                                                               |
   |                                                               |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Prefix #n Len  |                  Reserved                     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //                    Optional TLVs                            //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 15: Peer Prefix Advertisement Object Body Format for IPv6

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

No. of Prefix: 1 Byte. Identifies the number of prefixes that are advertised to the peer in the PPA object.

Reserved: 3 Bytes. MUST be set to zero while sending and MUST be ignored on receipt.

IPv4 Prefix: 4 Bytes. Identifies the prefix that will be sent to the peer identified by Peer IPv4 Address.

Prefix Len: 1 Byte. Identifies the length of the prefix.

Optional TLVs: TLVs that associated with this object, can be used to convey other necessary information for prefixes advertisement. No TLVs are currently defined.

For IPv6:

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

IPv6 Prefix: Identifies the prefix that will be sent to the peer identified by Peer IPv6 Address.

8. 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 the errors related to the newly defined objects that are related to Native IP TE procedures.

Error-Type	Meaning	Error-value
33	Native IP TE failure	
		0:Unassigned
		1:Local IP is in use
		2:Remote IP is in use
		3:Explicit Peer Route Error
		4:EPR/BPI Peer Info mismatch
		5:BPI/PPA Address Family mismatch
		6:PPA/BPI Peer Info mismatch
6	Mandatory Object missing	
		19:Native IP object missing
10	Reception of an invalid object	
		39:PCECC NATIVE-IP-TE-CAPABILITY bit is not set
19	Invalid Operation	
		22:Only one BPI,EPR or PPA object can be included in this message
		TBD1:Attempted Native-IP operations when capability was not advertised
		TBD2:Unknown Native-IP Info

Figure 16: Newly defined Error-Type and Error-Value

9. BGP Considerations

This document defines the procedures and objects to create the BGP sessions and advertise the associated prefixes dynamically. Only the key information, for example peer IP addresses, peer AS number are exchanged via the PCEP protocol. Other parameters that are needed for the BGP session setup should be derived from their default values.

When the PCE sends out the PCInitiate message with BPI object embedded to establish the BGP session between the PCC peers, PCC should report the BGP session status. For instance, the PCC could respond with "BGP Session Establishment In Progress" initially and on session establishment send another PCRpt message with state updated to "BGP Session Established". If there is any error during the BGP session establishment, the PCC should indicate the reason with the appropriate status value set in the BPI object.

Upon receiving such key information, the BGP module on the PCC should try to accomplish the task appointed by the PCEP protocol and report the successful status to the PCEP modules after the session is setup.

There is no influence on current implementation of BGP Finite State Machine (FSM). The PCEP focuses only on the success and failure status of BGP session, and acts upon such information accordingly.

The error handling procedures related to incorrect BGP parameters are specified in Section 6.1, Section 6.2, and Section 6.3.

10. Deployment Considerations

The information transferred in this document is mainly used for the 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 document.

The communication of PCE and PCC described in this document SHOULD follow the state synchronization procedures described in [RFC8232] , treat the three newly defined objects (BPI, EPR, PPA) associated with the same symbolic path name as the attribute of the same path in the LSP-DB (LSP State Database).

When PCE detects one or some of the PCCs are out of its control, it should recompute and redeploy the traffic engineering path for native IP on the currently active PCCs. The PCE should assure the avoidance of possible transient loop in such node failure when it deploys the explicit peer route on the PCCs.

In case of a PCE failure, a new PCE can gain control over the central controller instructions as described in [RFC9050].

As per the PCEP procedures in [RFC8281], the State Timeout Interval timer is used to ensure that a PCE failure does not result in automatic and immediate disruption for the services. Similarly, as per [RFC9050], the central controller instructions are not removed immediately upon PCE failure. Instead, they could be re-delegated to the new PCE before the expiration of this timer, or be cleaned up on the expiration of this timer. This allows for network clean up without manual intervention. The PCC supports the removal of CCI as one of the behaviors applied on expiration of the State Timeout Interval timer.

11. Manageability Considerations

11.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow the PCECC Native-IP capability to be enabled/disabled as part of the global configuration.

11.2. Information and Data Models

[RFC7420] describes the PCEP MIB; this MIB could be extended to get the PCECC Native-IP capability status. The PCEP YANG [I-D.ietf-pce-pcep-yang] module could be extended to enable/disable the PCECC Native-IP capability.

11.3. Liveness Detection and Monitoring

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

11.4. Verify Correct Operations

Verification of the mechanisms defined in this document can be built on those already listed in [RFC5440], [RFC8231] and [RFC9050]. Further, the operator needs to be able to verify the status of BGP sessions and prefix advertisements.

11.5. Requirements on Other Protocols

Mechanisms defined in this document requires the interaction with BGP. Section 9 describes in detail the considerations regarding to the BGP. During BGP session establishment, implementation MUST NOT allow the use local/remote IP address already sent in the BPI object.

11.6. Impact on Network Operations

[RFC8821] describes the various deployment considerations in CCDR architecture and their impact on network operations.

12. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to RFC 7942.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

12.1. Proof of Concept based on ODL

At the time of posting the -26 version of this document, there are no known implementations of this mechanism. A proof of concept for the overall design has been verified using another SBI protocol on the Open DayLight (ODL) controller.

12.2. ZTE

ZTE is preparing an implementation of this document as the time of posting the -29 version of this document.

13. Security Considerations

In this setup, the BGP sessions, prefix advertisement, and explicit peer route establishment are all controlled by the PCE. See [RFC4271] and [RFC4272] for BGP security considerations. Security considerations in [RFC5440], [RFC8231] and [RFC8281] should be considered. To prevent a bogus PCE from sending harmful messages to the network nodes, the network devices should authenticate the validity of the PCE and ensure a secure communication channel between them. Thus, the mechanisms described in [RFC8253] and [RFC9050] should be used.

14. IANA Considerations

14.1. 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 sub-registry, as follows:

Value	Description	Reference
4	Native IP TE Path	This document

14.2. PCECC-CAPABILITY sub-TLV’s Flag field

[RFC9050] 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:

Bit	Name	Reference
30	NATIVE IP	This document

14.3. PCEP Object

IANA is requested to allocate new codepoints in the "PCEP Objects" sub-registry as follows:

Object-Class Value	Name	Reference
44	CCI Object Object-Type 2: Native IP	This document
46	BGP Peer Info Object-Type 1: IPv4 address 2: IPv6 address	This document
47	Explicit Peer Route Object-Type 1: IPv4 address 2: IPv6 address	This document
48	Peer Prefix Advertisement Object-Type 1: IPv4 address 2: IPv6 address	This document

14.4. 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	Meaning	Error-value
6	Mandatory Object missing	19:Native IP object missing
10	Reception of an invalid object	39:PCECC NATIVE-IP-TE-CAPABILITY bit is not set
19	Invalid Operation	22:Only one BPI,EPR or PPA object can be included in this message TBD1:Attempted Native-IP operations when capability was not advertised TBD2:Unknown Native-IP Info
33	Native IP TE failure	1:Local IP is in use 2:Remote IP is in use 3:Explicit Peer Route Error 4:EPR/BPI Peer Info mismatch 5:BPI/PPA Address Family mismatch 6:PPA/BPI Peer Info mismatch

The reference for new Error-type/value should be set to this document.

14.5. CCI Object Flag Field

IANA is requested to create a new subregistry to manage the Flag field of the new CCI Object called "CCI Object Flag Field for Native-IP". New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

bit number (counting from bit 0 as the most significant bit)

capability description

defining RFC

Currently no flags are assigned.

14.6. BPI Object Status Code

IANA is requested to create a new sub-registry "BPI Object Status Code Field" within the "Path Computation Element Protocol (PCEP) Numbers". New values are assigned by Standards Action [RFC8126]. Each value should be tracked with the following qualities: value, meaning, and defining RFC. The following values are defined in this document:

Value	Meaning	Reference
0	Reserved	This document
1	BGP Session Established	This document
2	BGP Session Establishment In Progress	This document
3	BGP Session Down	This document
4-255	Unassigned	This document

14.7. BPI Object Error Code

IANA is requested to create a new sub-registry "BPI Object Error Code Field" within the "Path Computation Element Protocol (PCEP) Numbers". New values are assigned by Standards Action [RFC8126]. Each value should be tracked with the following qualities: value, meaning, and defining RFC. The following values are defined in this document:

Value	Meaning	Reference
0	Reserved	This document
1	ASes does not match, BGP Session Failure	This document
2	Peer IP can’t be reached, BGP Session Failure	This document
3-255	Unassigned	This document

14.8. BPI Object Flag Field

IANA is requested to create a new sub-registry "BPI Object Flag Field" within the "Path Computation Element Protocol (PCEP) Numbers". New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

bit number (counting from bit 0 as the most significant bit)

capability description

defining RFC

The following values are defined in this document:

Bit	Meaning	Reference
0-6	Unassigned	
7	T (IPnIP) bit	This document

15. Contributor

Dhruv Dhody has contributed to this document.

16. Acknowledgement

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

17. References

17.1. 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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

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

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

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

[RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <https://www.rfc-editor.org/info/rfc8231>.

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

[RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <https://www.rfc-editor.org/info/rfc8253>.

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

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

[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>.

17.2. Informative References

[I-D.ietf-pce-pcep-yang] Dhody, D., Beeram, V. P., Hardwick, J., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-yang-22, 11 September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-pcep-yang-22>.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <https://www.rfc-editor.org/info/rfc4272>.

[RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <https://www.rfc-editor.org/info/rfc7942>.

[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>.

[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>.

[RFC8821] Wang, A., Khasanov, B., Zhao, Q., and H. Chen, "PCE-Based Traffic Engineering (TE) in Native IP Networks", RFC 8821, DOI 10.17487/RFC8821, April 2021, <https://www.rfc-editor.org/info/rfc8821>.

Authors’ Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing
Beijing, 102209
China
Email: wangaijun@tsinghua.org.cn

Boris Khasanov
Yandex LLC
Ulitsa Lva Tolstogo 16
Moscow
Email: bhassanov@yahoo.com

Sheng Fang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
China
Email: fsheng@huawei.com

Ren Tan
Huawei Technologies
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