Internet Engineering Task Force (IETF)
Request for Comments: 9757
Category: Experimental
ISSN: 2070-1721

A. Wang
China Telecom
B. Khasanov
MTS Web Services (MWS)
S. Fang
Huawei Technologies
C. Zhu
ZTE Corporation

March 2025

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

Abstract

This document introduces extensions to the Path Computation Element Communication Protocol (PCEP) to support path computation in Native IP networks through a PCE-based central control mechanism known as Centralized Control Dynamic Routing (CCDR). These extensions empower a PCE to calculate and manage paths specifically for Native IP networks, thereby expanding PCEP's capabilities beyond its past use in MPLS and GMPLS networks. By implementing these extensions, IP network resources can be utilized more efficiently, facilitating the deployment of traffic engineering in Native IP environments.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for examination, experimental implementation, and evaluation.

This document defines an Experimental Protocol for the Internet community. This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9757.

Table of Contents

1. Introduction

Generally, Multiprotocol Label Switching Traffic Engineering (MPLS-TE) requires the corresponding network devices to support the Resource ReSerVation Protocol (RSVP) [RFC3209] and the Label Distribution Protocol (LDP) [RFC5036] to ensure 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 a solution based on multiple Border Gateway Protocol (BGP) sessions. It requires only the PCE to send the instructions to the Path Computation Clients (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 the BGP peer, peer prefix advertisement, and 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.

2.1. Use of RBNF

The message formats in this document are illustrated using Routing Backus-Naur Form (RBNF) encoding, as specified in [RFC5511]. The use of RBNF is illustrative only and may elide certain important details; the normative specification of messages is found in the prose description. If there is any divergence between the RBNF and the prose, the prose is considered authoritative.

2.2. Experimental Status Consideration

The procedures outlined in this document are experimental. The experiment aims to explore the use of PCE (and PCEP) for E2E traffic assurance in Native IP networks through multiple BGP sessions. Additional implementation is necessary to gain a deeper understanding of the operational impact, scalability, and stability of the mechanism described. Feedback from deployments will be crucial in determining whether this specification should advance from Experimental to the IETF Standards Track.

3. Terminology

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

Additionally, the following terminology is used in this document:

BPI: BGP Peer Info

CCDR: Centralized Control Dynamic Routing

CCI: Central Controller Instructions (defined in [RFC9050])

E2E: End-to-End

EPR: Explicit Peer Route

Native IP network: Network that forwards traffic based solely on the IP address, instead of another indicator, for example, MPLS, etc.

PCECC: PCE as a Central Controller (defined in [RFC8283])

PPA: Peer Prefix Advertisement

PST: Path Setup Type (defined in [RFC8408])

SRP: Stateful PCE Request Parameter (defined in [RFC8231])

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 the PCEP speakers' PCECC capability. A new flag is defined in the PCECC-CAPABILITY sub-TLV for Native IP:

N (NATIVE-IP-TE-CAPABILITY - 1 bit - 30): When set to 1 by a PCEP speaker, this flag indicates that the PCEP speaker is capable of TE in a Native IP network, as specified in this document. Both the PCC and PCE MUST set this flag to support this extension.

If a PCEP speaker receives the PATH-SETUP-TYPE-CAPABILITY TLV with the newly defined PST, 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) and

* terminate the PCEP session.

If a PCEP speaker receives the PATH-SETUP-TYPE-CAPABILITY TLV with the newly defined PST, 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) and

* terminate the PCEP session.

If one or both speakers (PCE and PCC) have not indicated the support 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 on the OPEN messages, the receiver of the message MUST:

* send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=29 (Attempted Native IP operations when the capability was not advertised) and

* terminate the PCEP session.

5. PCEP Messages

The 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 establish multiple BGP sessions, deploy the E2E Native IP TE path, and advertise route prefixes 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 the Central Controller Instructions (CCI). [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), Explicit Peer Route (EPR), and Peer Prefix Advertisement (PPA)) 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 RFC 5440

<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 the PCInitiate message is used for Native IP instructions, i.e., when the CCI Object-Type is 2, the SRP, LSP, and CCI objects MUST be present. Error handling for missing SRP, LSP, or CCI objects MUST be performed as specified in [RFC9050]. Additionally, exactly one object among the BPI, EPR, or PPA objects MUST be present. The PCEP-specific LSP identifier (PLSP-ID) and Symbolic Path Name TLVs 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, or PPA objects are identified by the same Symbolic Path Name.

If none of the BPI, EPR, or PPA objects 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 is more than one 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).

When the PCInitiate message is not used for Native IP instructions,

i.e., when the CCI Object-Type is not equal to 2, the BPI, EPR, and PPA objects SHOULD NOT be present. If present, they MUST be ignored by the receiver.

To clean up 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].

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

The error handling for missing CCI objects is as per [RFC9050]. Furthermore, one and only one BPI, EPR, or PPA object MUST be present.

If none of the BPI, EPR, or PPA objects 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 is more than one 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).

When the PCInitiate message is not used for Native IP instructions, i.e., when the CCI Object-Type is not equal to 2, the BPI, EPR, and PPA objects SHOULD NOT be present. If present, they MUST be ignored by the receiver.

6. PCECC Native IP TE Procedures

The detailed procedures for the TE in the 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 are exchanged between a PCE and each BGP peer (acting as the PCC), which needs to establish a 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 is sent to the BGP router and/or RR (which are acting as the PCC).

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 an RR. The PCInitiate message is sent to the BGP routers R1, R3, and R7, which need to establish a BGP session.

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

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

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

When the PCC receives this message with the R bit set to 1 in the SRP object in the PCInitiate message, the PCC MUST clear the BGP configuration and tear down the BGP session that is indicated by the BPI object.

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

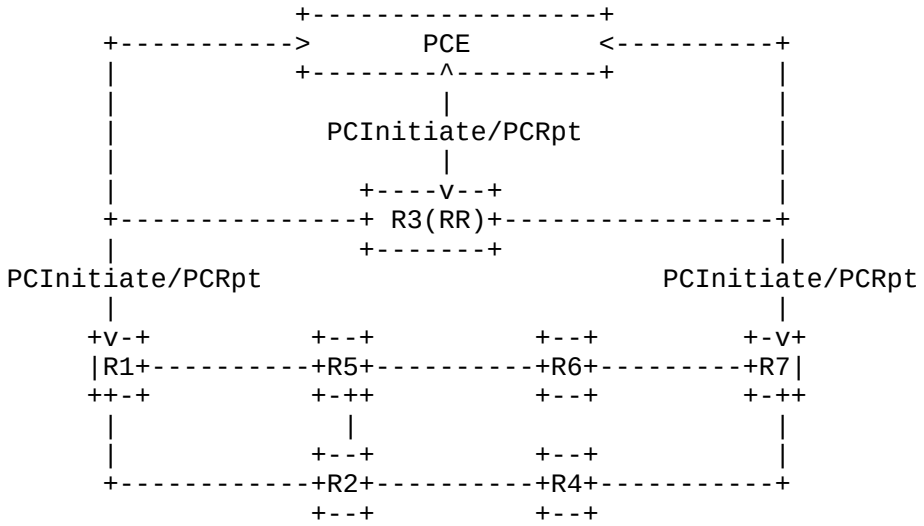


Figure 1: BGP Session Establishment Procedures (R3 acts as the RR)

The message peers, message types, message key parameters, and procedures in the above figure are shown below:

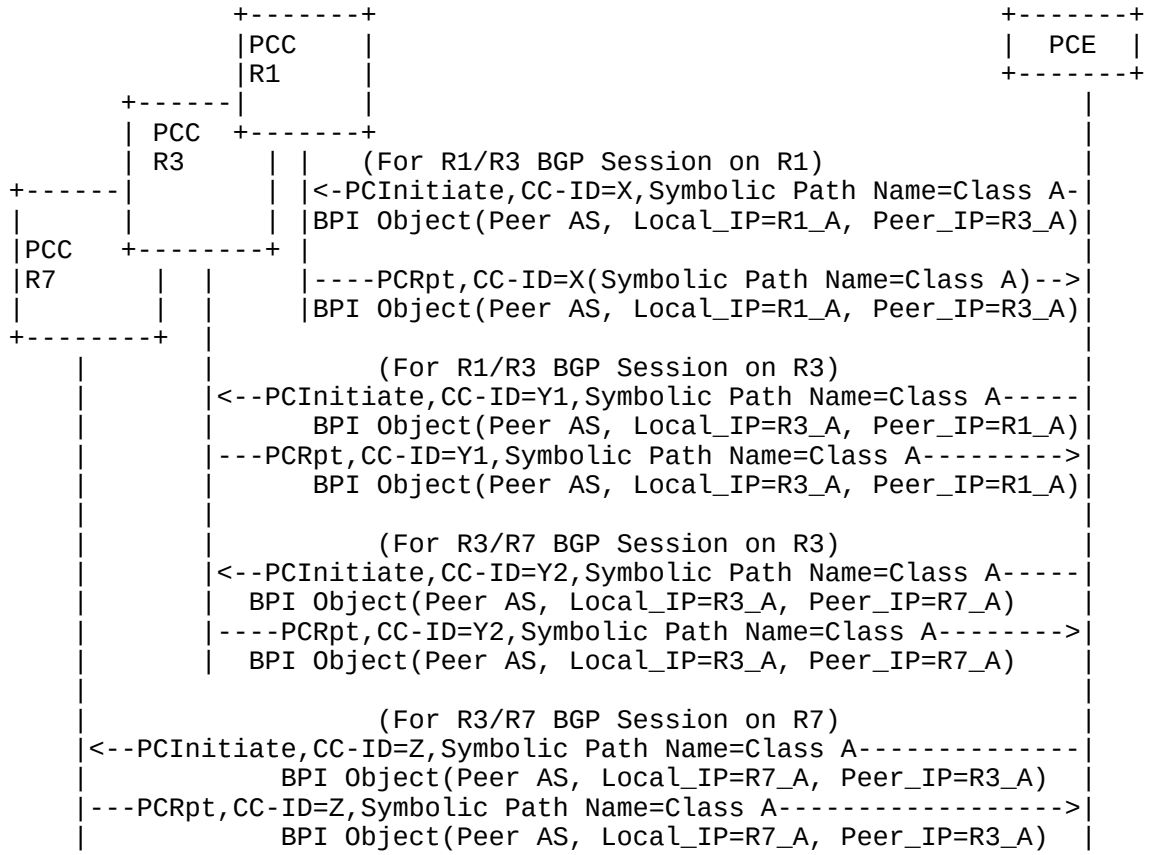


Figure 2: Message Information and Procedures

The Local/Peer IP Address MUST be dedicated to the usage of the Native IP TE solution and MUST NOT be used by other BGP sessions that are established manually or in other ways. If the Local IP Address or Peer IP Address within the BPI object is used in other existing BGP sessions, the PCC MUST report such an 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 MUST report such information via a PCRpt message with the status field set to "BGP session down" in the associated BPI object. The error code field within the BPI object SHOULD indicate the reason that leads to the BGP session being down. In the future, when the BGP session is up again, the PCC MUST report that as well via the PCRpt message with the status field set to "BGP Session Established".

6.2. Explicit Route Establishment Procedures

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

For the purpose of explicit route addition, the PCInitiate message ought to be sent to every router on the explicit path. In the example, for the explicit route from R1 to R7, the PCInitiate message is sent to R1, R2, and R4, as shown in Figure 3. For the explicit route from R7 to R1, the PCInitiate message is sent to R7, R4, and R2, as shown in Figure 5.

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

When the PCC successfully installs the explicit route to the peer, it MUST report the result via the PCRpt message, with the EPR object and the corresponding SRP and CCI objects included.

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

When the PCC has removed the explicit route that is indicated by this object, it MUST report the result via the PCRpt message, with the EPR object and the corresponding SRP and CCI objects included.

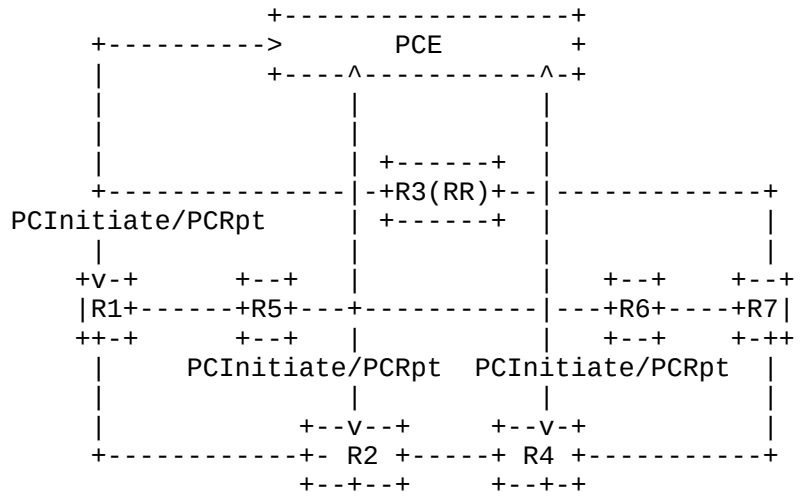
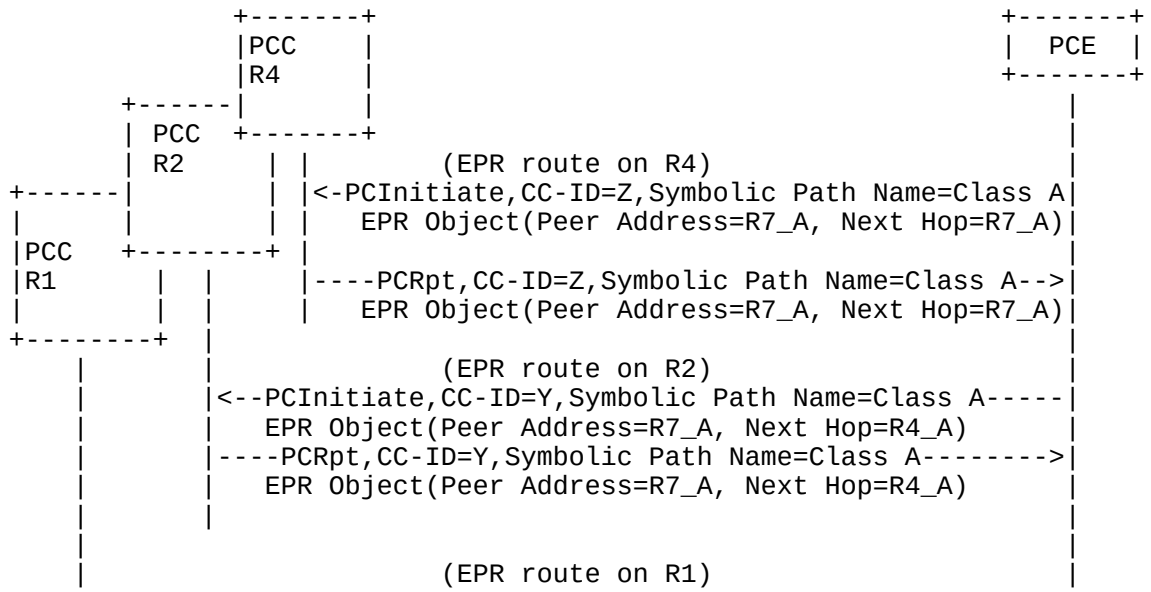


Figure 3: Explicit Route Establish Procedures (from R1 to R7)

The message peers, message types, message key parameters, and procedures in the above figure are shown below:



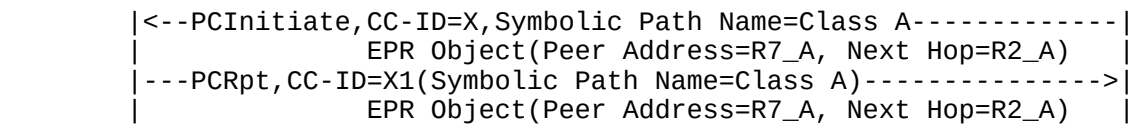


Figure 4: Message Information and Procedures

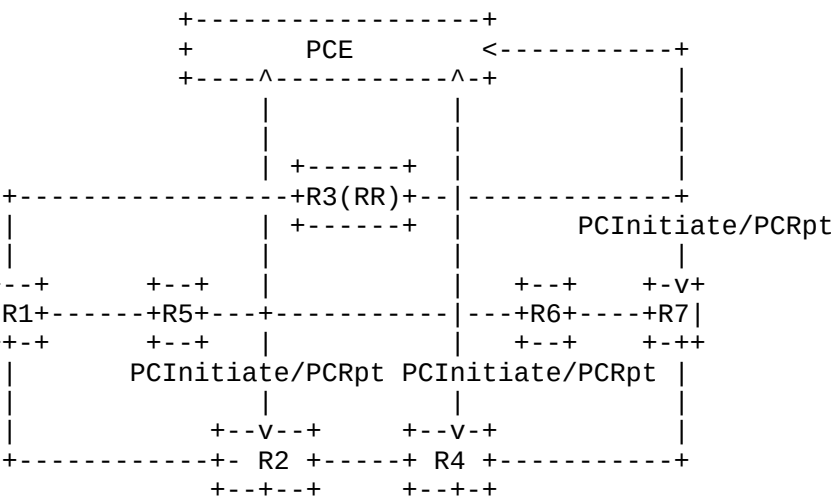


Figure 5: Explicit Route Establish Procedures (from R7 to R1)

The message peers, message types, message key parameters, and procedures in the above figure are shown below:

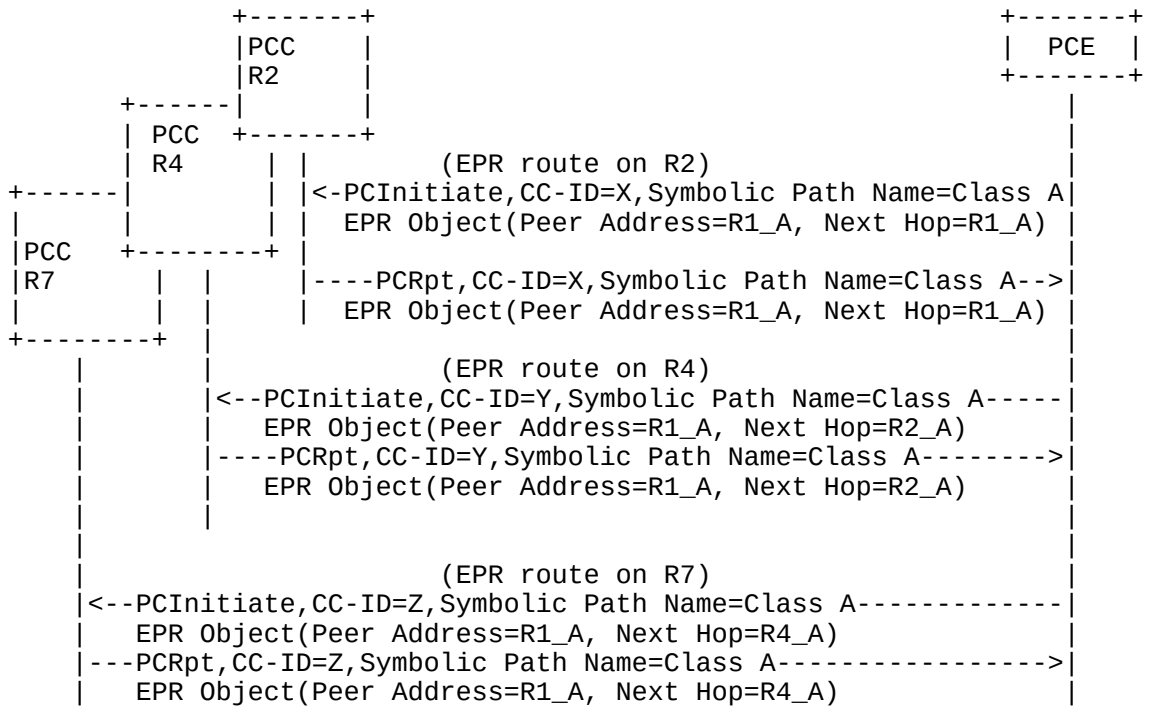


Figure 6: Explicit Route Establish Procedures (from R7 to R1)

To avoid the transient loop while deploying the explicit peer route, the EPR object MUST be sent to the PCCs in the reverse order of the E2E path. To remove the explicit peer route, the EPR object MUST be sent to the PCCs in the same order as 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 a different next-hop address.

The PCC MUST verify that the next-hop address is reachable. In case of failure, the PCC MUST send the corresponding error via a PCErr message, with the error information: Error-Type=33 (Native IP TE failure) and 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 the PCC for the same path that is identified by Symbolic Path Name TLV, a PCErr message MUST be reported, with the error information Error-Type=33 (Native IP TE failure) and 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 MUST first instruct the new CCI with the updated EPR corresponding to the new next hop to use and then instruct the removal of the older CCI.

6.3. BGP Prefix Advertisement Procedures

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

The PCInitiate message SHOULD be sent to the PCC that acts as a BGP peer edge router only. In the example, it is sent to R1 and R7, respectively.

When the PCC receives the PPA and the CCI object (with the R bit set to 0 in the SRP object) in the PCInitiate message, the PCC SHOULD send the prefixes indicated in this object to the identified BGP peer via the corresponding BGP session [RFC4271].

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

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

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

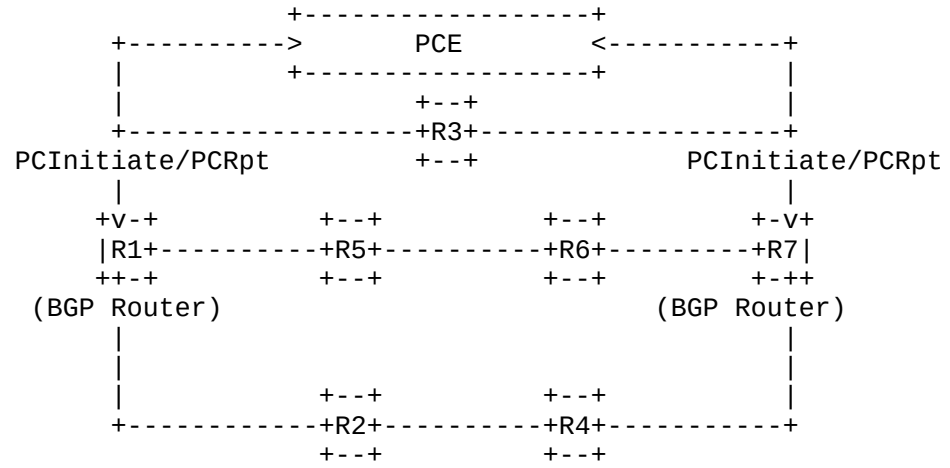


Figure 7: BGP Prefix Advertisement Procedures

The message peers, message types, message key parameters, and procedures in the above figure are shown below:

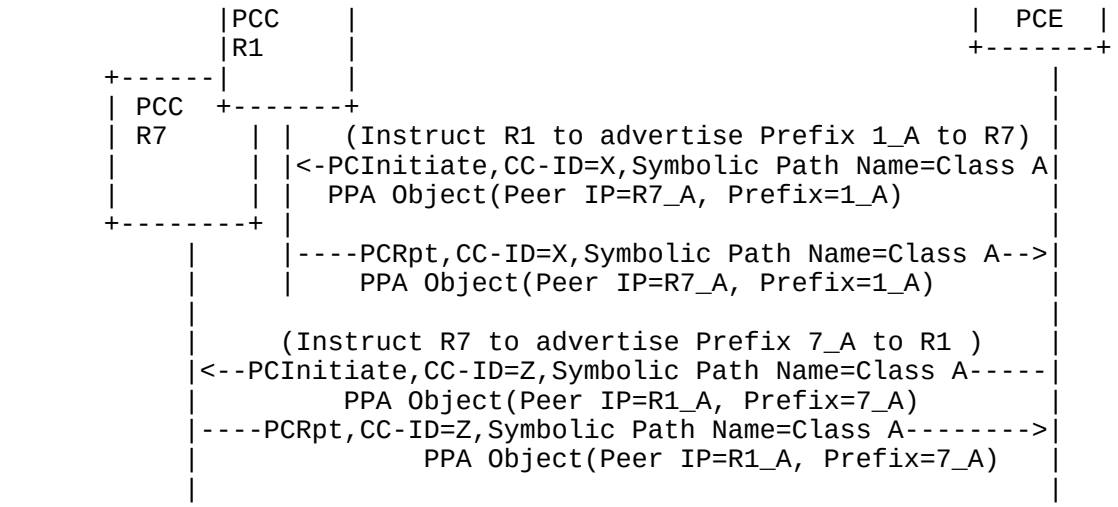


Figure 8: Message Information and Procedures

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

When the peer info is not the same as the peer info that is indicated in the BPI object in the PCC for the same path that is identified by Symbolic Path Name TLV, an error, i.e., Error-Type=33 (Native IP TE failure) and Error-value=6 (PPA/BPI Peer Info mismatch), MUST 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 the 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 traffic is coming into the network from different attached points but to the same destination, 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 ensure that traffic enters the network via R1 and exits the network at R7 via R5-R6-R7. If some traffic enters the network via the R2 router, passes through R5, and exits at R7, they may share the priority path among R5-R6-R7, which may not be the desired effect.

The above normal traffic forwarding behavior is clarified as a Raw mode forwarding strategy. Such a mode can only achieve the moderate traffic path control effect. To achieve the strict traffic path control effect, the entry point MUST 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 established via Section 6.1. Such forwarding behavior is called the Tunnel mode forwarding strategy. For simplicity, the IP-in-IP tunnel type [RFC2003] is used between the BGP peers by default.

The selection of Raw mode and Tunnel mode forwarding strategies are controlled via the T bit in the BPI object, which is defined in Section 7.2

6.5. Cleanup

To remove the Native IP state from the PCC, the PCE MUST send explicit CCI cleanup instructions for PPA, EPR, and BPI objects,

respectively, with the R bit 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=30 (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, redelegation, and cleanup 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.

The CCI Object-Type is 2 for Native IP, as follows:

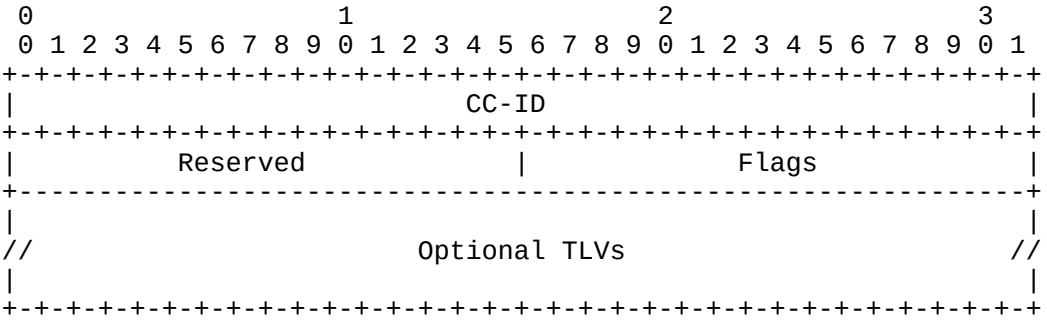


Figure 9: CCI Object for Native IP

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

Reserved: 2 bytes. Set to zero while sending and ignored on receipt.

Flags: 2 bytes. Used to carry any additional information about the Native IP CCI. Currently, no flag bits are defined. Unassigned flags are set to zero while sending and ignored on receipt.

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

7.2. BGP Peer Info Object

The BGP Peer Info (BPI) object is used to specify the information about the peer with which the PCC wants to establish the BGP session. This object is 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 is sent to the source router, RR, and destination router, respectively.

By default, the Local/Peer IP Address MUST be a unicast address and dedicated to the usage of the Native IP TE solution and MUST NOT be

used by other BGP sessions that are established by manual or other configuration mechanisms.

The BGP Peer Info Object-Class is 46.

The 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:

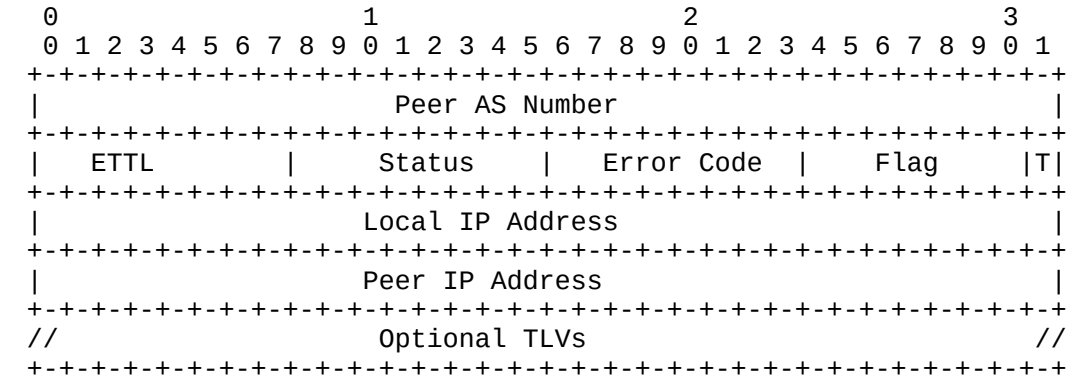


Figure 10: BGP Peer Info Object Body Format for IPv4

The format of the BGP Peer Info object body for IPv6 (Object-Type=2) is as follows:

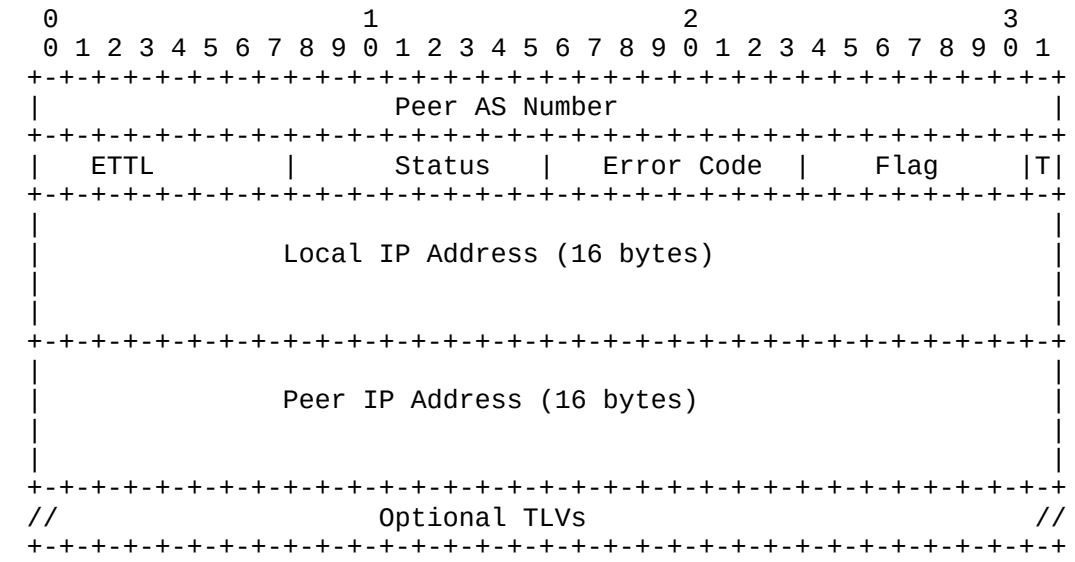


Figure 11: BGP Peer Info Object Body Format for IPv6

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

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

Status: 1 byte. Indicates the BGP session status between the peers. Its 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. Indicates the reason that the BGP session can't be established.

0: Unspecific

1: ASes do 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 the T bit is set, the traffic SHOULD be sent in the IP-in-IP tunnel (the tunnel source is the Local IP Address, and the tunnel destination is the Peer IP Address). When the T bit is cleared, the traffic is sent via its original source and destination address. The Tunnel mode (i.e., the T bit is set) is used when the operator wants to ensure only the traffic from the specified (entry, exit) pair, and the Raw mode (i.e., the T bit is clear) is used when the operator wants to ensure 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): Unicast IP address of the local router, used to peer with another end router. When the Object-Type is 1, the length is 4 bytes; when the Object-Type is 2, the length is 16 bytes.

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

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

When the PCC receives a BPI object, with Object-Type=1, it SHOULD try to establish a BGP session with the peer in AFI/SAFI=1/1.

When the PCC receives a BPI object, with Object-Type=2, it SHOULD try to establish a BGP session with the peer in AFI/SAFI=2/1.

7.3. Explicit Peer Route Object

The Explicit Peer Route (EPR) 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 ought to be sent to all the devices on the path that are calculated by the PCE. Although the object is named "Explicit Peer Route", it can be seen that the routes it installs are simply host routes. The use of this object to install host routes for any purpose other than reaching the corresponding peer address on each device that is on the E2E Native IP TE path is outside the scope of this specification.

By default, the path established by this object MUST have higher priority than the other paths calculated by the dynamic IGP protocol

and MUST have lower priority than the static route configured by manual, NETCONF, or any other static means.

The Explicit Peer Route Object-Class is 47.

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

The format of the Explicit Peer Route object body for IPv4 (Object-Type=1) is as follows:

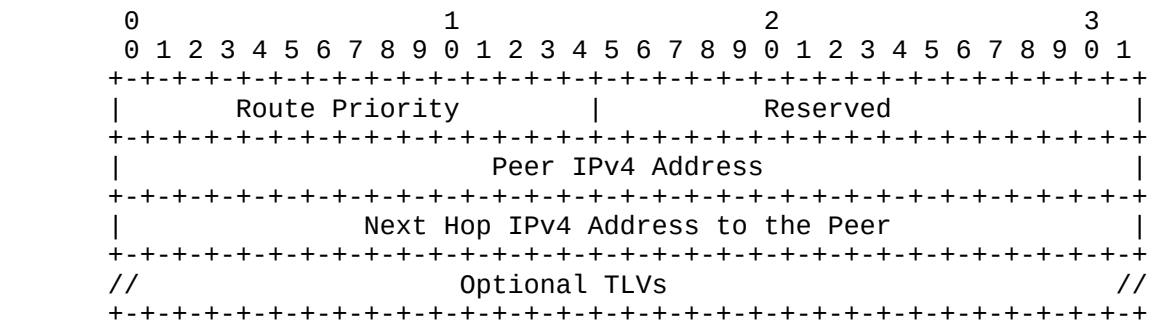


Figure 12: Explicit Peer Route Object Body Format for IPv4

The format of the Explicit Peer Route object body for IPv6 (Object-Type=2) is as follows:

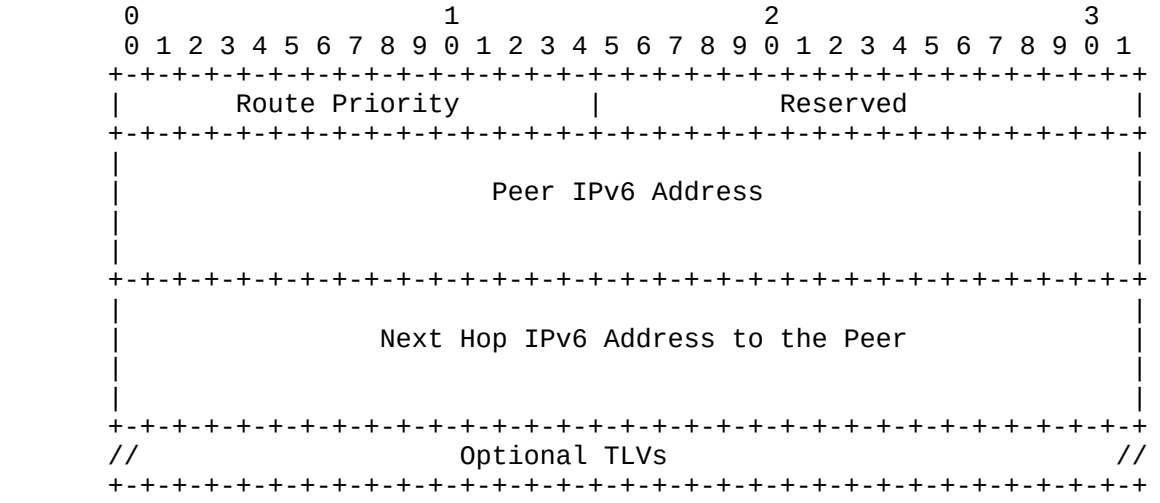


Figure 13: Explicit Peer Route Object Body Format for IPv6

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: Set to zero while sending and ignored on receipt.

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

Next Hop (IPv4/IPv6) Address to the Peer: Indicates the next-hop address (4/16 bytes) to the corresponding peer address.

Optional TLVs: TLVs that are 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 (PPA) object is defined to specify the

IP prefixes that are advertised to the corresponding peer. This object only needs to be included and sent to the source/destination router of the E2E path.

The prefix information included in this object MUST only be advertised to the indicated peer and SHOULD NOT be advertised to other BGP peers.

The Peer Prefix Advertisement Object-Class is 48.

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

The format of the Peer Prefix Advertisement object body for IPv4 is as follows:

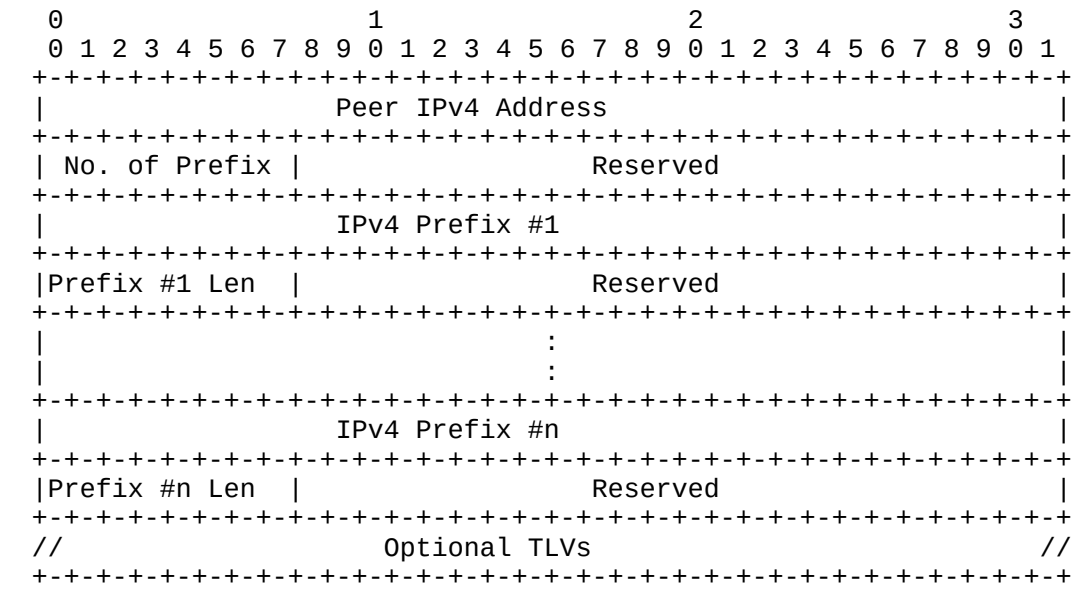
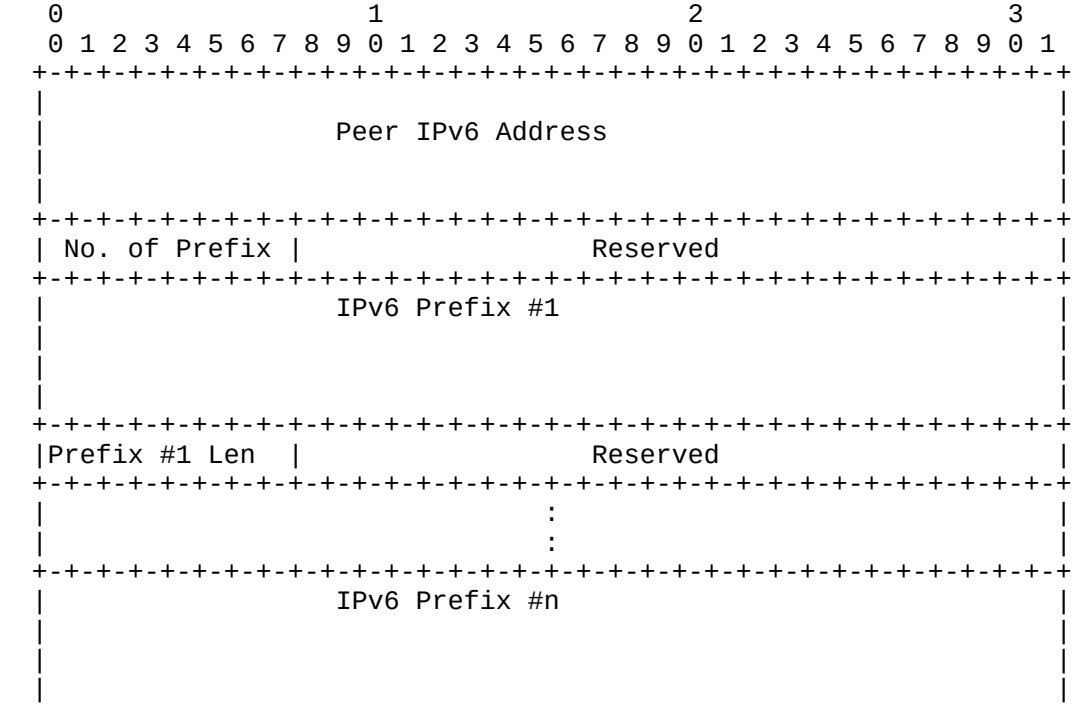


Figure 14: Peer Prefix Advertisement Object Body Format for IPv4

The format of the Peer Prefix Advertisement object body for IPv6 is as follows:



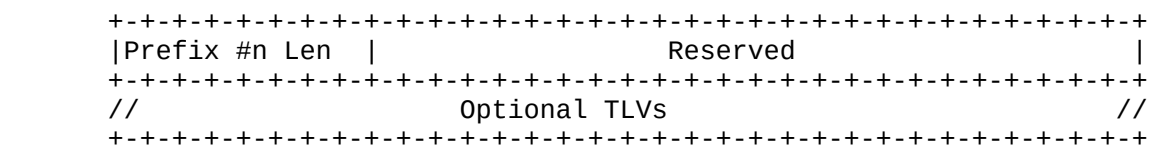


Figure 15: Peer Prefix Advertisement Object Body Format for IPv6

Common Fields:

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

Reserved: 3 bytes. Ought to be set to zero while sending and ignored on receipt.

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

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

For IPv4:

Peer IPv4 Address: 4 bytes. Identifies the Peer IPv4 Address that the associated prefixes will be sent to.

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

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 the Peer IPv6 Address.

If in the future a requirement is identified to advertise IPv4 prefixes towards an IPv6 peering address or IPv6 prefixes towards an IPv4 peering address, then a new Peer Prefix Advertisement Object-Type can be defined for these purposes.

8. New Error-Type 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. See Table 4 for the newly defined Error-Type and Error-values.

9. BGP Considerations

This document defines procedures and objects to create the BGP sessions and to advertise the associated prefixes dynamically. Only the key information, for example, Peer IP Addresses, and Peer AS numbers are exchanged via the PCEP. 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 the BPI object embedded to establish the BGP session between the PCC peers, the 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 the 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 and report the successful status to the PCEP modules after the session is set up.

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

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

10. Deployment Considerations

The information transferred in this document is mainly used for the BGP session setup, explicit route deployment, and prefix distribution. The planning, allocation, and distribution of the peer addresses within IGP need to be accomplished in advance, and they are out of the scope of this document.

The communication of PCE and PCC described in this document MUST follow the State Synchronization procedures described in [RFC8232], i.e., treat the three newly defined objects (BPI, EPR, and PPA) associated with the same Symbolic Path Name as the attribute of the same path in the LSP Database (LSP-DB).

When the PCE detects that one or some of the PCCs are out of its control, it MUST recompute and redeploy the traffic engineering path for Native IP on the currently active PCCs. The PCE MUST ensure the avoidance of the 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 redelegated to the new PCE before the expiration of this timer or be cleaned up on the expiration of this timer. This allows for network cleanup without manual intervention. The PCC supports the removal of CCI as one of the behaviors applied on the 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 module [YANG-PCEP] 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 beyond 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 require the interaction with BGP. Section 9 describes in detail the considerations regarding the BGP. During the BGP session establishment, the Local/Peer IP Address MUST be dedicated to the usage of the Native IP TE solution and MUST NOT be used by other BGP sessions that are established manually or in other ways.

11.6. Impact on Network Operations

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

12. Security Considerations

In this setup, the BGP sessions, prefix advertisement, and explicit peer route establishment are all controlled by the PCE. See [RFC4271] for classical BGP implementation security considerations and [RFC4272] for classical BGP vulnerabilities analysis. Security considerations in [RFC5440] for the basic PCEP, [RFC8231] for PCEP extension for stateful PCE, and [RFC8281] for PCE-initiated LSP setup SHOULD be considered. To prevent a bogus PCE from sending harmful messages to the network nodes, the network devices SHOULD authenticate the PCE and ensure a secure communication channel between them. Thus, the mechanisms described in [RFC8253] for the usage of TLS for PCEP and [RFC9050] for protection against malicious PCEs SHOULD be used.

If the default values discussed in Section 9 aren’t enough and securing the BGP transport is required (for example, by using TCP Authentication Option (TCP-AO) [RFC5925]), a suitable value can be provided through the addition of optional TLVs to the BGP Peer Info object that conveys the necessary additional information (for example, a key chain [RFC8177] name).

13. IANA Considerations

13.1. PCEP Path Setup Types

[RFC8408] created the "PCEP Path Setup Types" registry within the "Path Computation Element Protocol (PCEP) Numbers" registry group. IANA has allocated a new codepoint within this registry, as follows:

Value	Description	Reference
4	Native IP TE Path	RFC 9757

Table 1: PCEP Path Setup Types Registry

13.2. PCECC-CAPABILITY Sub-TLV Flag Field

[RFC9050] created the "PCECC-CAPABILITY sub-TLV" registry within the "Path Computation Element Protocol (PCEP) Numbers" registry group to

manage the value of the PCECC-CAPABILITY sub-TLV's 32-bit Flag field. IANA has allocated a new bit position within this registry, as follows:

Bit	Name	Reference
30	Native IP	RFC 9757

Table 2: PCECC-CAPABILITY Sub-TLV Registry

13.3. PCEP Objects

IANA has allocated new codepoints in the "PCEP Objects" registry, as follows:

Object-Class Value	Name	Object-Type	Reference
44	CCI Object-Type	2: Native IP	RFC 9757
46	BGP Peer Info Object-Type	0: Reserved	RFC 9757
		1: IPv4 address	RFC 9757
		2: IPv6 address	RFC 9757
47	Explicit Peer Route Object-Type	0: Reserved	RFC 9757
		1: IPv4 address	RFC 9757
		2: IPv6 address	RFC 9757
48	Peer Prefix Advertisement Object-Type	0: Reserved	RFC 9757
		1: IPv4 address	RFC 9757
		2: IPv6 address	RFC 9757

Table 3: PCEP Objects Registry

13.4. PCEP-Error Objects

IANA has allocated a new Error-Type and several Error-values in the "PCEP-ERROR Object Error Types and Values" registry within the "Path Computation Element Protocol (PCEP) Numbers" registry group, as follows:

Error-Type	Meaning	Error-value	Reference

6	Mandatory Object missing	19: Native IP object missing	RFC 9757
10	Reception of an invalid object	39: PCECC NATIVE-IP-TE-CAPABILITY bit is not set	RFC 9757
19	Invalid Operation	22: Only one BPI, EPR, or PPA object can be included in this message	RFC 9757
		29: Attempted Native IP operations when the capability was not advertised	RFC 9757
		30: Unknown Native IP Info	RFC 9757
33	Native IP TE failure	0: Unassigned	RFC 9757
		1: Local IP is in use	RFC9757
		2: Remote IP is in use	RFC 9757
		3: Explicit Peer Route Error	RFC 9757
		4: EPR/BPI Peer Info mismatch	RFC 9757
		5: BPI/PPA Address Family mismatch	RFC 9757
		6: PPA/BPI Peer Info mismatch	RFC 9757

Table 4: PCEP-ERROR Object Error Types and Values Registry

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

13.5. CCI Object Flag Field

IANA has created the "CCI Object Flag Field for Native IP" registry to manage the 16-bit Flag field of the new CCI object. New values are to be assigned by IETF Review [RFC8126]. Each bit should be tracked with the following qualities:

* bit number (counting from bit 0 as the most significant bit and bit 15 as the least significant bit)

* capability description

* defining RFC

Currently, no flags are assigned.

13.6. BPI Object Status Codes

IANA has created the "BPI Object Status Code Field" registry within the "Path Computation Element Protocol (PCEP) Numbers" registry group. New values are assigned by IETF Review [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	RFC 9757
1	BGP Session Established	RFC 9757
2	BGP Session Establishment In Progress	RFC 9757
3	BGP Session Down	RFC 9757
4-255	Unassigned	RFC 9757

Table 5: BPI Object Status Code Field Registry

13.7. BPI Object Error Codes

IANA has created the "BPI Object Error Code Field" registry within the "Path Computation Element Protocol (PCEP) Numbers" registry group. New values are assigned by IETF Review [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	RFC 9757
1	ASes do not match - BGP Session Failure	RFC 9757
2	Peer IP can’t be reached - BGP Session Failure	RFC 9757
3-255	Unassigned	RFC 9757

Table 6: BPI Object Error Code Field Registry

13.8. BPI Object Flag Field

IANA has created the "BPI Object Flag Field" registry within the "Path Computation Element Protocol (PCEP) Numbers" registry group. New values are to be assigned by IETF Review [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 (IP-in-IP) bit	RFC 9757

Table 7: BPI Object Flag Field

Registry

14. References

14.1. Normative References

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, DOI 10.17487/RFC2003, October 1996, <https://www.rfc-editor.org/info/rfc2003>.

[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>.

[RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <https://www.rfc-editor.org/info/rfc5511>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <https://www.rfc-editor.org/info/rfc5925>.

[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>.

14.2. Informative References

[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>.

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

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, DOI 10.17487/RFC5036, October 2007, <https://www.rfc-editor.org/info/rfc5036>.

[RFC8177] Lindem, A., Ed., Qu, Y., Yeung, D., Chen, I., and J. Zhang, "YANG Data Model for Key Chains", RFC 8177, DOI 10.17487/RFC8177, June 2017, <https://www.rfc-editor.org/info/rfc8177>.

[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>.

[YANG-PCEP] 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-30, 26 January 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-pcep-yang-30>.

Acknowledgements

Thanks to Mike Koldychev, Susan Hares, Siva Sivabalan, and Adam

Simpson for their valuable suggestions and comments.

Contributors

Ren Tan and Dhruv Dhody have contributed to this document.

Authors' Addresses

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

Boris Khasanov
MTS Web Services (MWS)
Andropova av., 18/9
Moscow
115432
Russian Federation
Email: bhassanov@yahoo.com

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

Chun Zhu
ZTE Corporation
50 Software Avenue, Yuhua District
Nanjing
Jiangsu, 210012
China
Email: zhu.chun1@zte.com.cn