Updates for PCEP Extension for Native IP Network

draft-ietf-pce-pcep-extension-native-ip-16

A. Wang (China Telecom)
B. Khasanov (Yandex)
Sheng Fang (Huawei Technologies)
Ren Tan(Huawei Technologies)
Chun Zhun(ZTE Corporation)

IETF Interim, August 2021

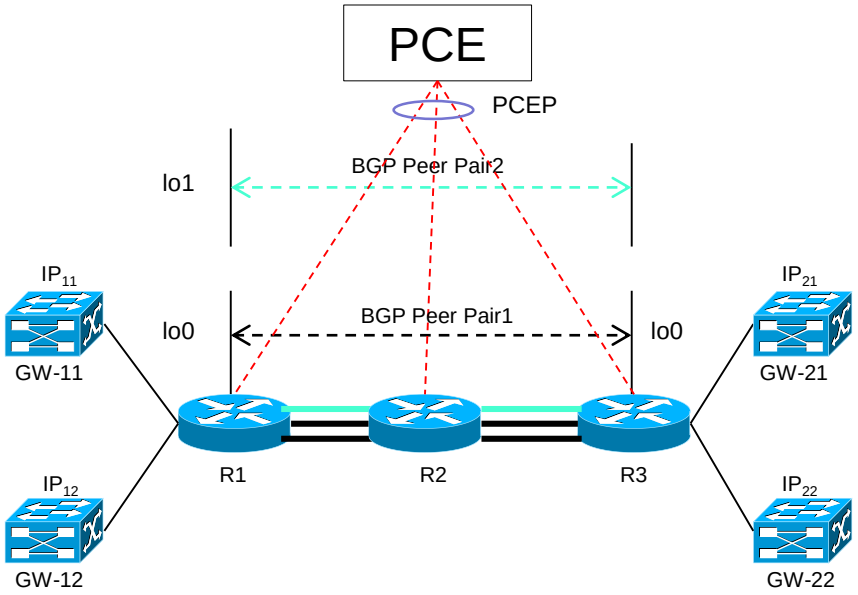
Motivation

- Introduce the updates for “PCEP extension for Native IP Network”
- Seek feedbacks for the overall solution from IDR experts
- Ready for WG Last Call

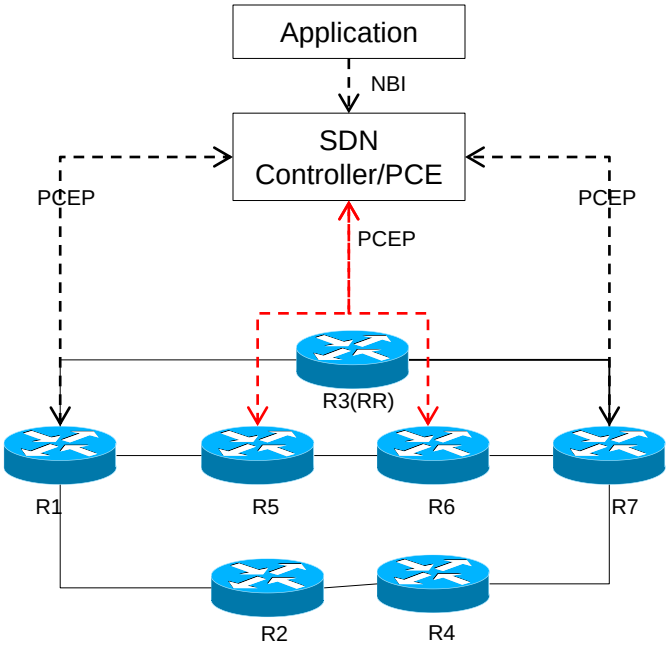
Backgrounds of the Solution

- RFC8735 describes the scenarios and simulation results for TE in Native IP network
- RFC 8821 describes the architecture for providing traffic engineering in a native IP network by using multiple BGP sessions and a PCE-based central control mechanism.
- This document describes the PCEP extensions and procedures to practically build such architecture.

Overview of the Solution



Dual/Multi-BGP Solution



Simplified CCDR* Architecture in a Large Network

- Building Dual/Multi BGP sessions between edge routers upon request via PCEP
- Advertises different prefixes via different BGP sessions, w/PCEP-based setup
- Steer traffic towards particular routes via BGP next-hop w/PCEP-based setup

CCDR* - Centralized Control Dynamic Routing, RFC 8735

PCEP Extensions

From Section 4-10 of the document:

- Capability Advertisement
- Related PCEP messages
- New PCEP Objects
- New Error-Type and Error-Value
- CCDR Native IP Procedures
- Operations Consideration

PCEP Capability Advertisement

- RFC8408 defines the Path Setup Type Capability TLV to indicate the path type supported by the PCE and PCC

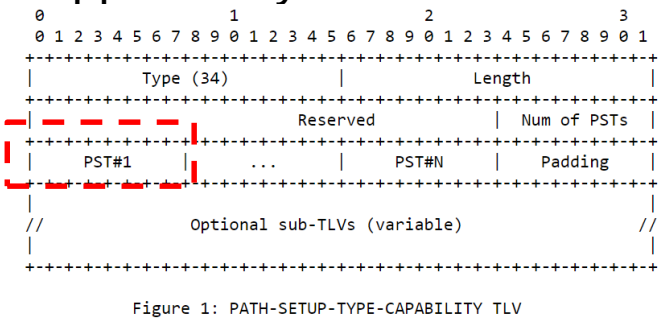


Figure 1: PATH-SETUP-TYPE-CAPABILITY TLV

  - New PST (TBD) is defined for Native IP path

- Draft-ietf-pce-pcep-extension-for-pce-controller defines the PCECC capability sub-tlv to exchange information about the PCECC capability

  - N(NATIVE-IP-TE-CAPABILITY-1 bit=TBD) is defined for PCEP speaker

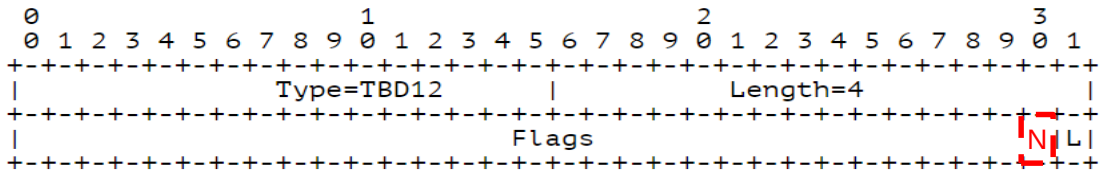


Figure 7: PCECC Capability sub-TLV

Updated PCEP Messages

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

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

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

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

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

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

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

When PCInitiate message is used create Native IP instructions, the SRP and CCI objects MUST be present. The error handling for missing SRP or CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. Further either one of BPI, EPR, or PPA object MUST be present. If none of them are present, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD (Native IP object missing).

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

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

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

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

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

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

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

The error handling for missing CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. Further either one of BPI, EPR, or PPA object MUST be present. If none of them are present, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD (Native IP object missing).

SRP – Stateful Request parameters Object RFC8231, CCI – Central Controller Instructions Object, BPI - BGP Peer Info Object, ERP – Explicit Peer Route Object, PPA - Peer Prefix Association Object.

New PCEP Objects(1/4)

CCI: Central Controller Instructions

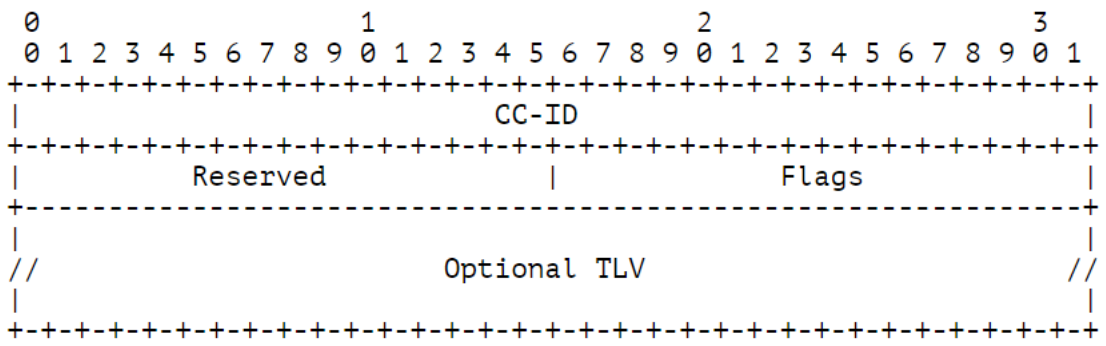


Figure 5: CCI Object for Native IP

CC-ID is described in draft-ietf-pce-pcep-extension-for-pce-controller

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

Symbolic Path Name TLV(RFC 8231) MUST be included in the above CCI-Object

New PCEP Objects(2/4)

BPI (BGP Peer Info) Object-Class is TBD

BPI Object-Type is for IPv4 and 2 for IPv6

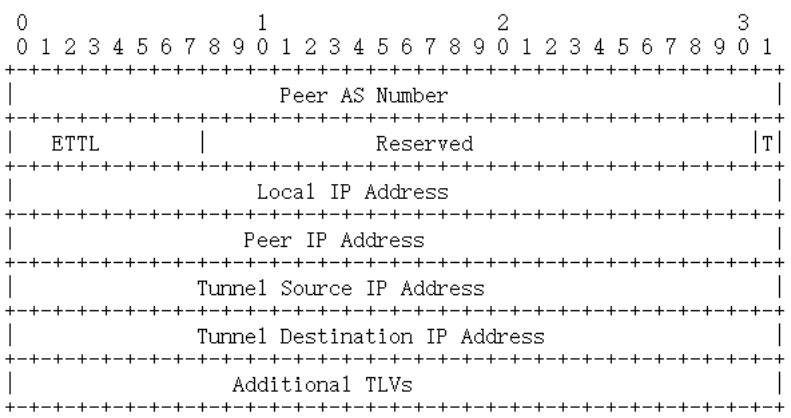


Figure 6: BGP Peer Info Object Body Format for IPv4

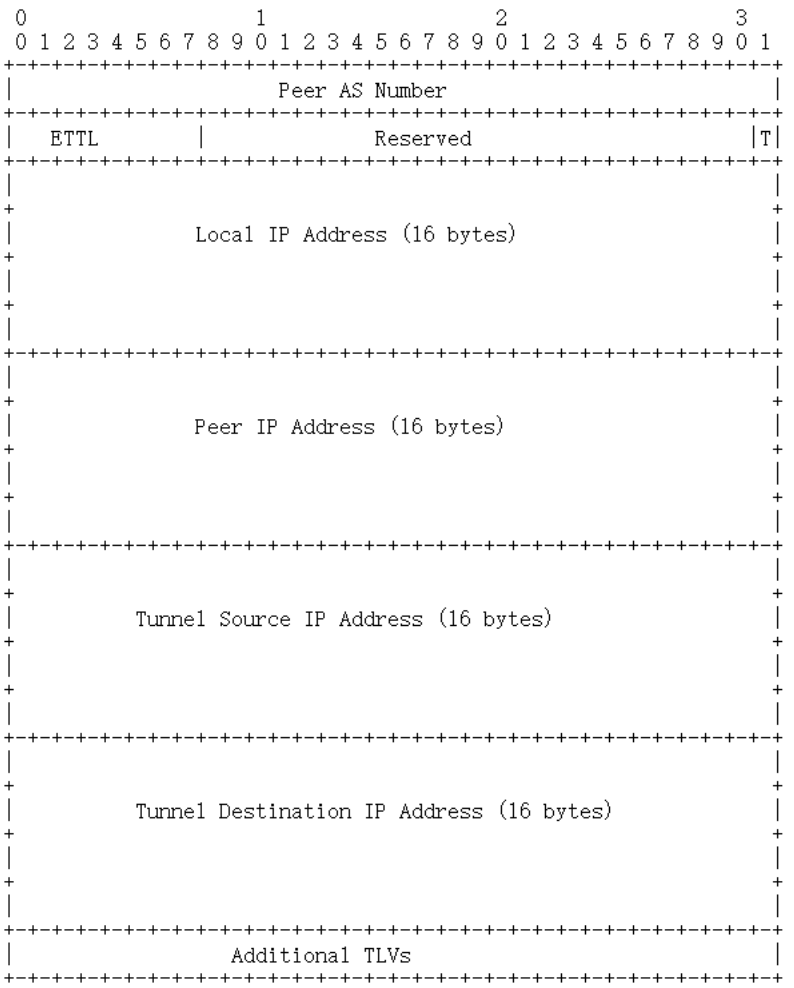


Figure 7: BGP Peer Info Object Body Format for IPv6

New PCEP Objects(3/4)

EPR (Explicit Peer Route) Object-Class is TBD

EPR 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:

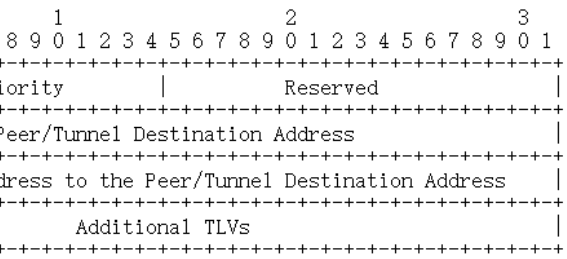


Figure 8: Explicit Peer Route Object Body Format for IPv4

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

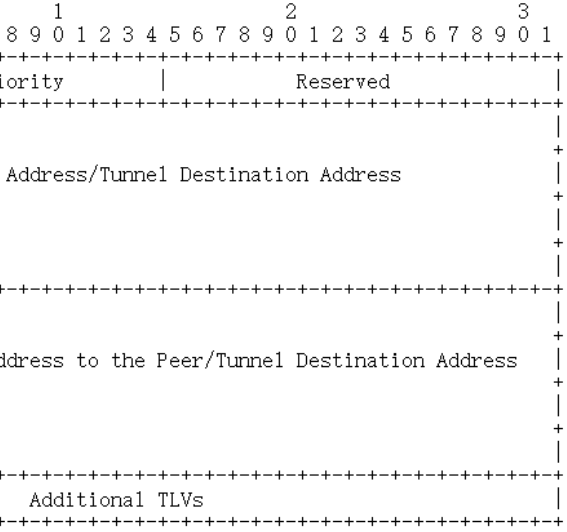


Figure 9: Explicit Peer Route Object Body Format for IPv6

New PCEP Objects(4/4)

PPA (Peer Prefix Association) Object-Class is TBD

PPA Object-Type is 1 for IPv4 and 2 for IPv6

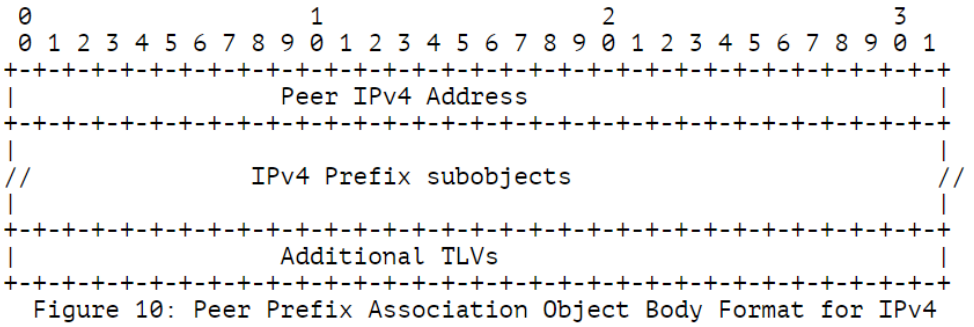


Figure 10: Peer Prefix Association Object Body Format for IPv4

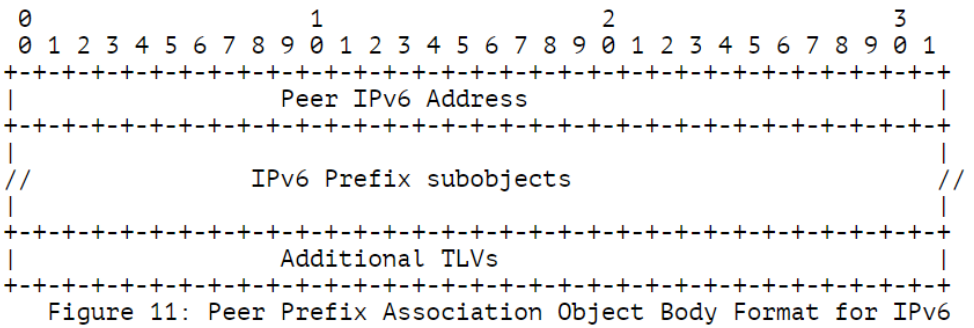


Figure 11: Peer Prefix Association Object Body Format for IPv6

IPv4 Prefix sub-object/IPv6 Prefix sub-object is defined in RFC3209

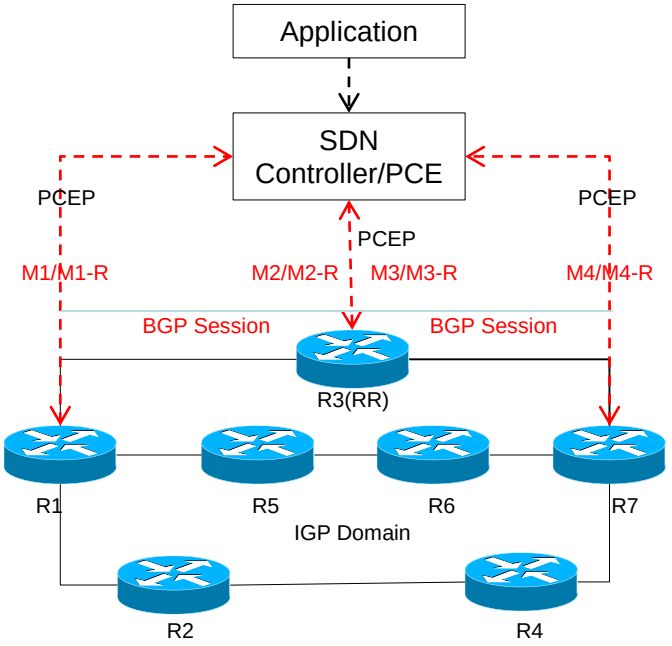
New Error-Type and Error-Values

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

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

CCDR Native IP Procedures(1/3)

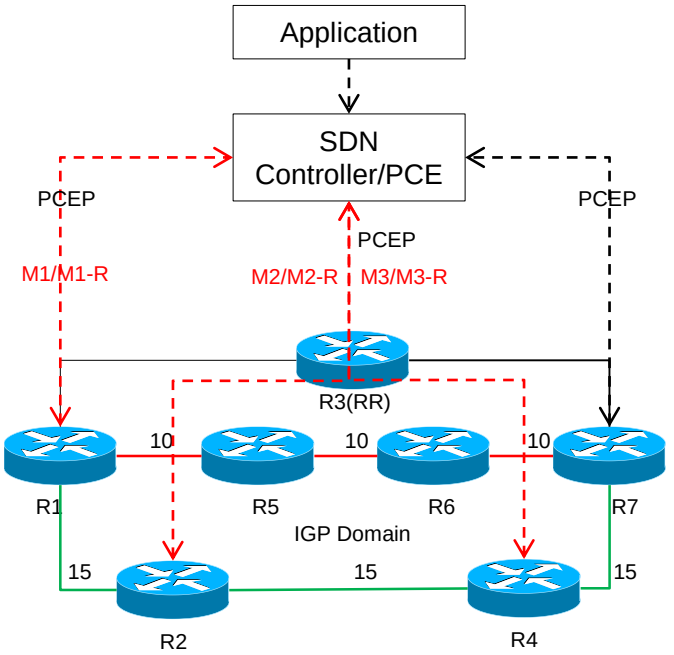
- PCEP messages (PCInitiate/PCRpt) are sent to R1、R3(RR)、R7 respectively, to build/remove the BGP sessions between R1/R3(RR) and R3(RR)/R7
- BPI Object is included
- “CC-ID” is unique, but “Symbolic Path Name” is constant for the E2E path
- “Local/Peer IP address” and “Peer AS” number is assigned within the BPI Object



BGP Session Establishment Procedures

CCDR Native IP Procedures(2/3)

- PCEP messages (PCInitiate/PCRpt) are sent to on-path routers R1、R2、R4 respectively, to install/remove the explicit route to the BGP nexthop
- EPR Object is included
- “CC-ID” is unique, but “Symbolic Path Name” is constant for the E2E path
- “Peer Address” and “Next Hop” information is assigned within the BPI Object. The route priority etc. for such path could also be assigned
- Reverse Path is built similarly from R7/R4/R2

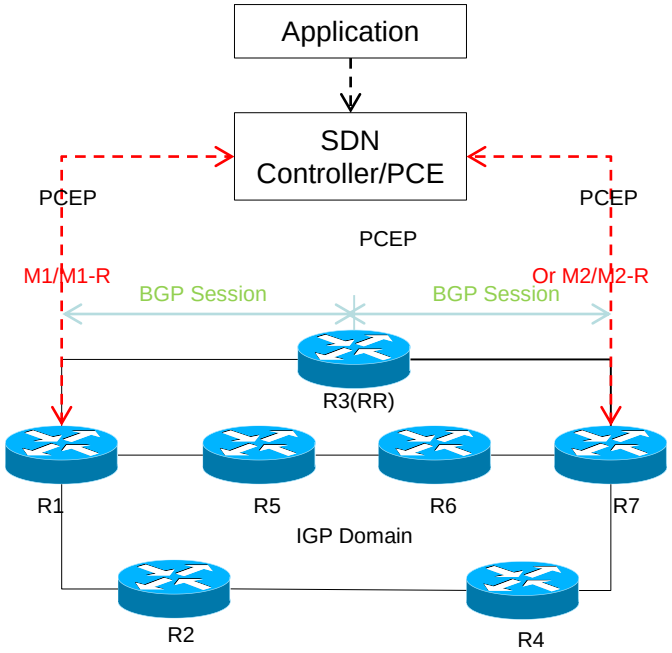


Explicit Route Establish Procedures (From R1 to R7))

Red Link represent congested path (IGP shortest path)
Green Link represent idle path (explicit route from PCE)

CCDR Native IP Procedures(3/3)

- PCEP messages (PCInitiate/PCRpt) are sent only to edge routers R1 or R7, to advertise/revoke the prefixes that associated with different BGP sessions
- PPA Object is included
- “CC-ID” is unique, but “Symbolic Path Name” is constant for the E2E path
- “Peer Address” and “Advertised Prefixes” information is assigned within the PPA Object
- Same AF prefixes should be advertised via the same AF sessions



BGP Prefix Advertisement Procedures

Operations Consideration

- The information transferred in this draft is mainly used for the light weight BGP session setup, explicit route deployment and the prefix distribution
- The planning, allocation and distribution of the peer addresses within IGP domain should be done in advance
- The state synchronization between PCE and PCC should follow the procedure that described in RFC8232
- When PCE detects one or some of PCCs are out of control, it should recompute and redeploy a traffic engineering path for native IP on active PCCs
- When PCC detects that connection with a PCE is lost, it should stale the information that initiated by PCE. A PCE should assure the avoidance of possible transient loop in such node failure case when it deploys the explicit peer route on the PCCs.

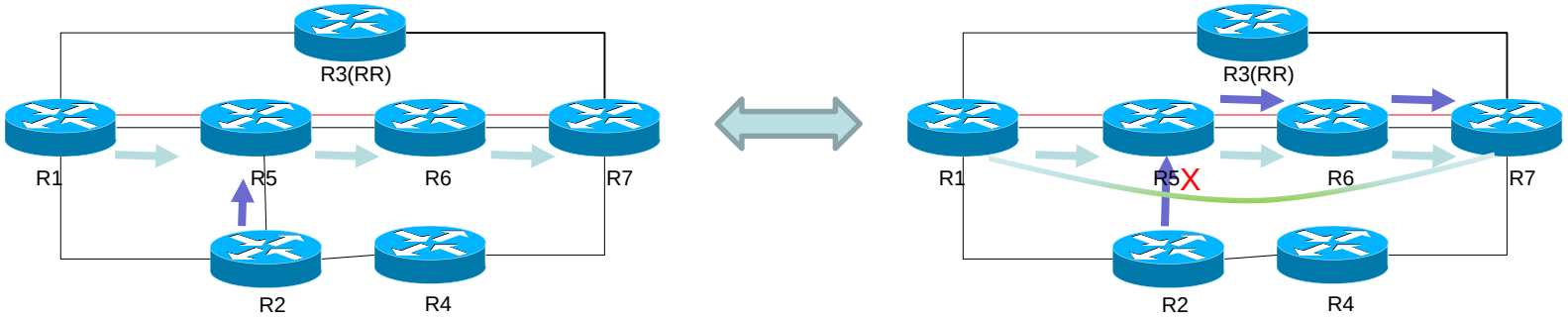
Next Step

1. Comments/Q&A
2. WG Last Call?

Aijun Wang@ChinaTelecom
Khasanov.Boris@Yandex
Sheng Fang@Huawei
Ren Tan@Huawei
C.Zhu@ZTE

IETF Interim(111&112) @Online

BPI & EPR Object Updates Considerations



Native Traffic Forwarding

Tunneled Traffic Forwarding

- ✓ Destination of user traffic based
- ✓ Traffic from different sources to the same destination may share the priority path
- ✓ Moderate traffic path control

- ✓ Destination of tunnel based
- ✓ Traffic for different (source, address) tuple are put into different tunnel
- ✓ Strict traffic path control