PCE as a central Controller (PCECC) Extensions

draft-zhao-pce-pcep-extension-for-pce-controller-08

draft-zhao-pce-pcep-extension-pce-controller-sr-03

Quintin Zhao
Zhenbin Li
Dhruv Dhody
Satish Karunanithi
Adrian Farrel
Chao Zhou

Introduction

RFC 8283 is published

- An architecture for use of PCE/PCEP in a network with central control.
- Introduces the architecture for PCE as a central controller and examines the motivations/applicability for PCEP as a control protocol in this environment.
- A PCE-based central controller can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it.

During IETF 101

Should PCEP have full-fledged SDN capabilities?

- Discussion on-list last year
- What did we conclude?
  - Using PCEP for SDN-like function is reasonable and has already happened
  - PCEP-SDN offers an alternative to other SBI protocols (e.g. netconf, OpenFlow)
  - PCEP-SDN is not “replacing” anyone’s control plane

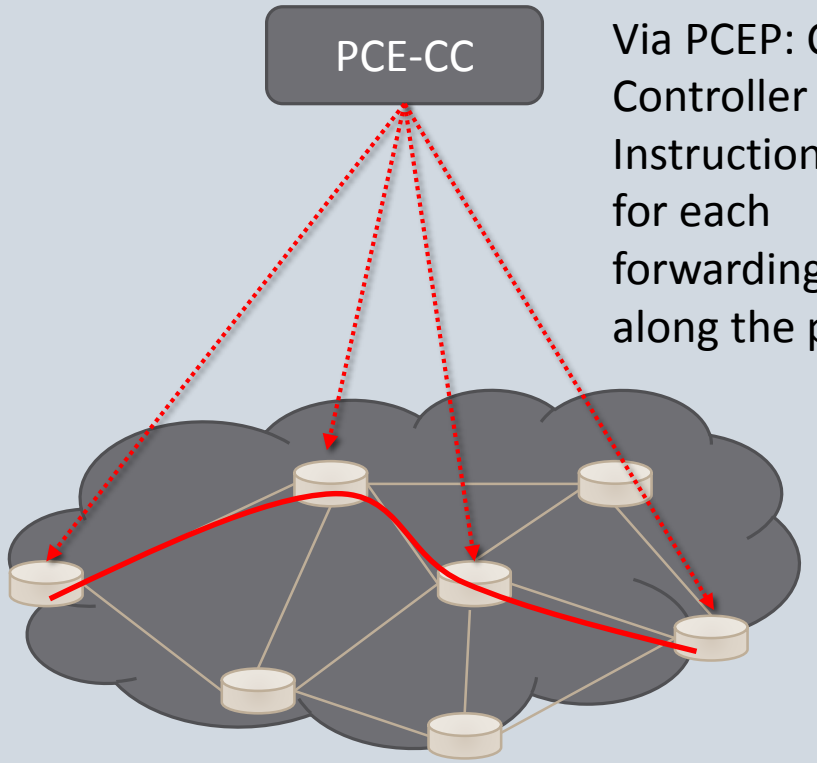
Basic PCECC Mode

LSPs are provisioned as explicit label instructions at each hop on the end-to-end path.

Each router along the path must be told what label forwarding instructions to program and what resources to reserve.

The controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

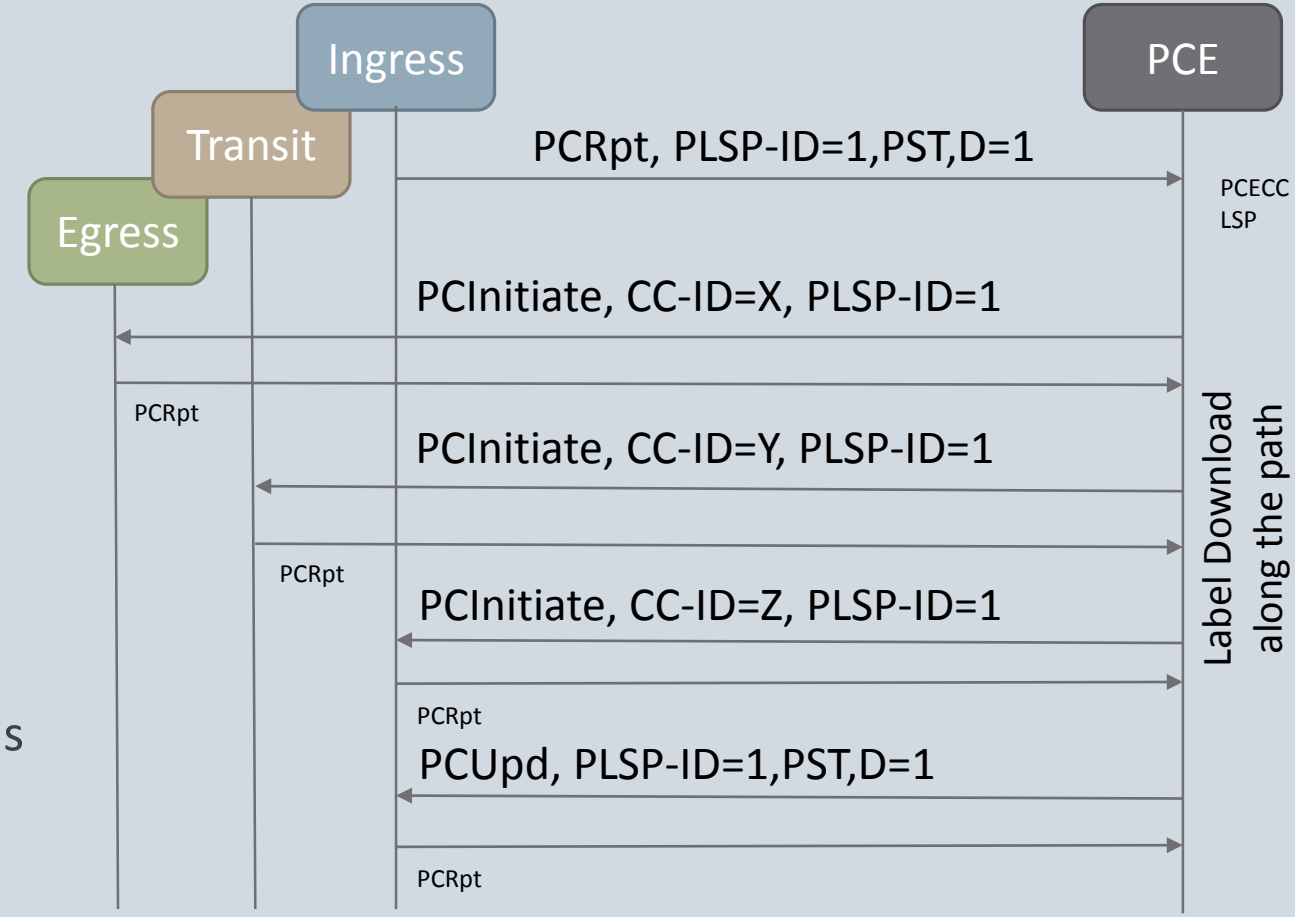
PCECC will take responsibility for managing some part of the MPLS label space for each of the routers that it controls



Via PCEP: Central Controller Instructions (CCI) for each forwarding action along the path

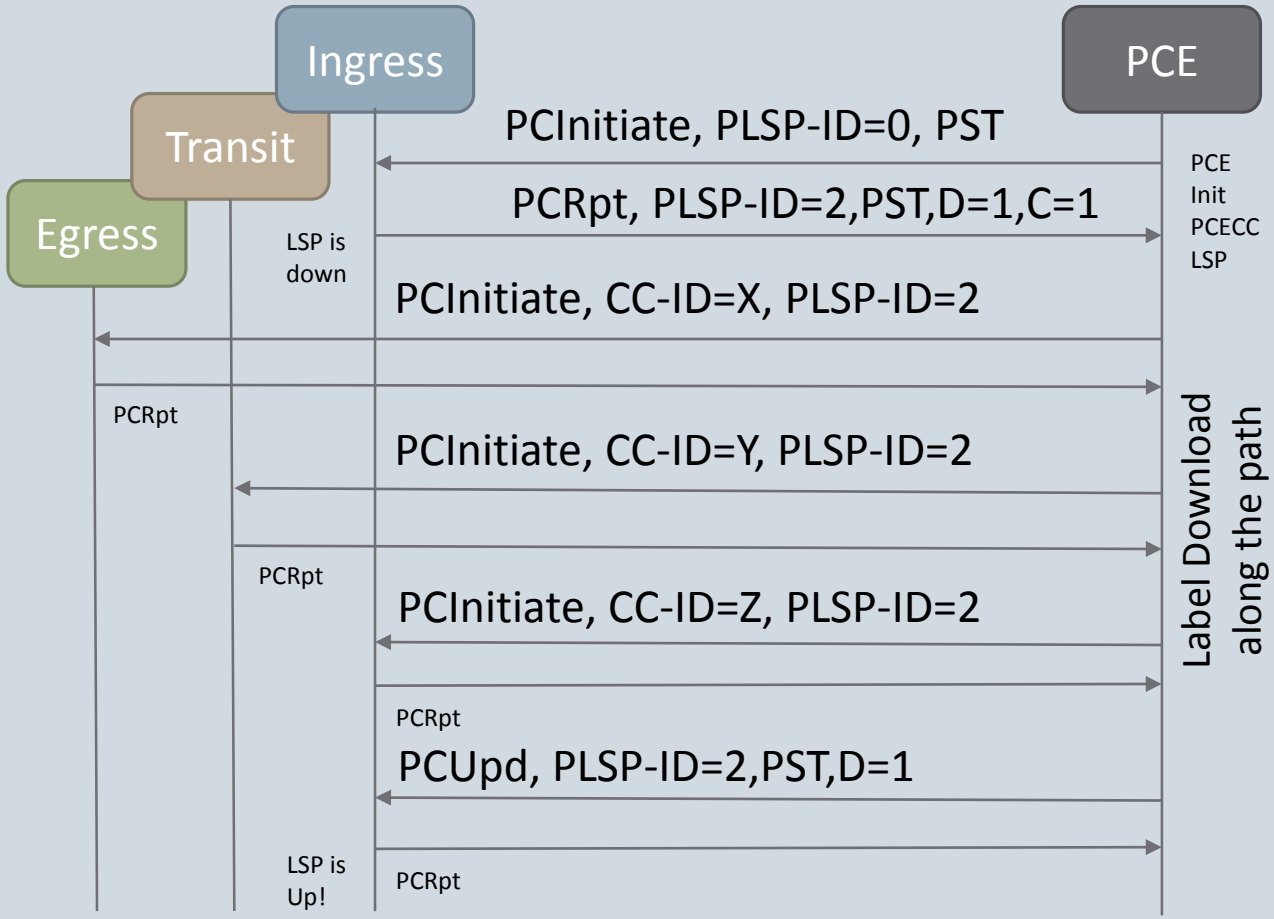
Central Controller's Instruction (CCI)

- For Label Download/Cleanup
- PCInitiate message is extended to include CCI operations
  - CC-ID uniquely identify the central controller instruction as assigned by the PCE
  - LSP object is included to identify the LSP for which this instruction is used
  - PLSP-ID as assigned by the ingress is used
- In response to the PCInitiate, PCC sends the PCRpt message (which is also extended to include CCI)
- PCRpt / PCUpd message to ingress PCC remains as per RFC 8231



CCI – PCE-Initiated LSP

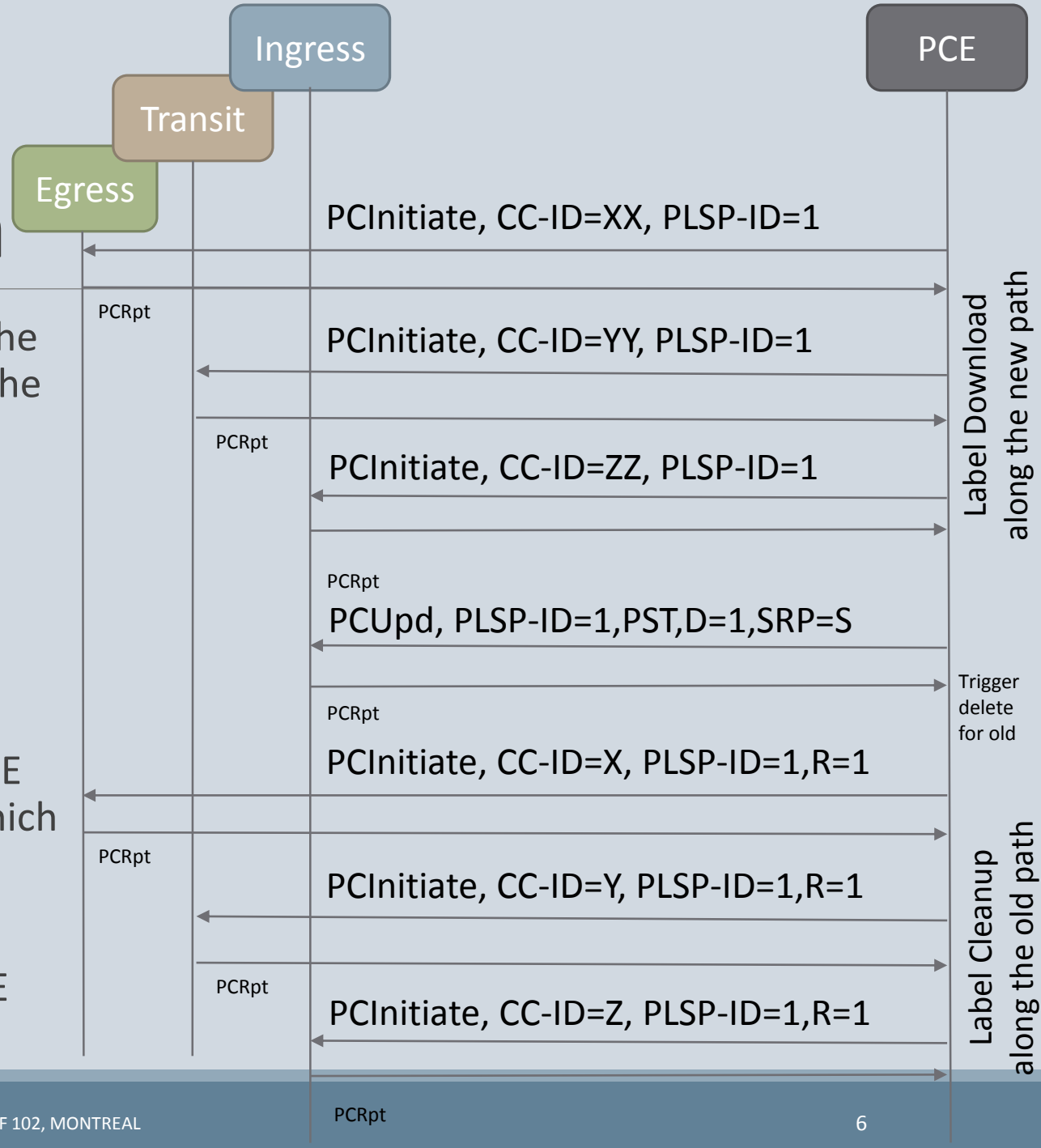
- Incase of PCE-Initiated LSP, the PST is set for PCECC and a PLSP-ID is assigned by the ingress.
- At this time the LSP is not up.
- Using the CCI in the PCInitiate message, the label download instructions are made along the path and the LSP object includes the identification details as assigned by the ingress.
- Once the labels are download, the PCECC marks the LSP as up by sending the PCUpd message to the ingress.
- *For cleanup the R flag is set in the SRP object in the PCInitiate message*



CCI – LSP Modification

Make-Before-Break!

- To follow the make-before-break procedures, the PCECC first update new instructions based on the updated path and then update to ingress to switch traffic, before cleaning up the old instructions.
- A new CC-ID is used to identify the updated instruction, the existing identifiers in the LSP object identify the existing LSP.
- Once new instructions are downloaded, the PCE further updates the new path at the ingress which triggers the traffic switch on the updated path.
- The Ingress PCC acknowledges with a PCRpt message, on receipt of PCRpt message, the PCE does cleanup operation for the old LSP.



SR PCECC Mode

PCECC can use PCEP for SR SID (Segment Identifier) distribution on the SR nodes.

- SR SID is just another central controller instruction (CCI)

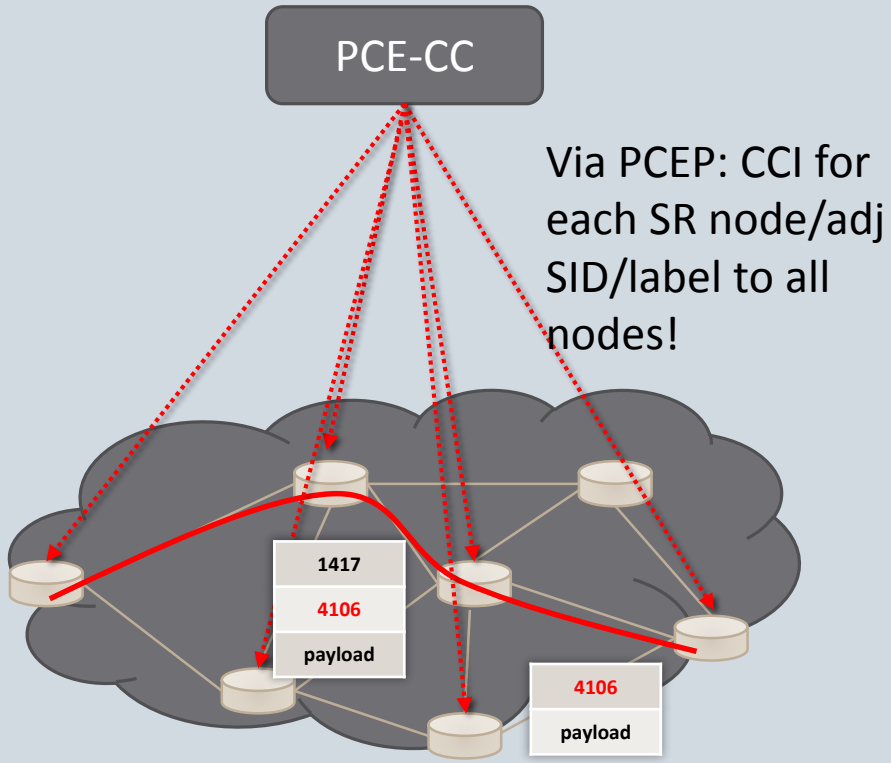
PCECC needs to be in control of label space to make SR SID allocation

- Node/Prefix
- Adjacency

Rest of the PCEP-SR procedures remains unchanged

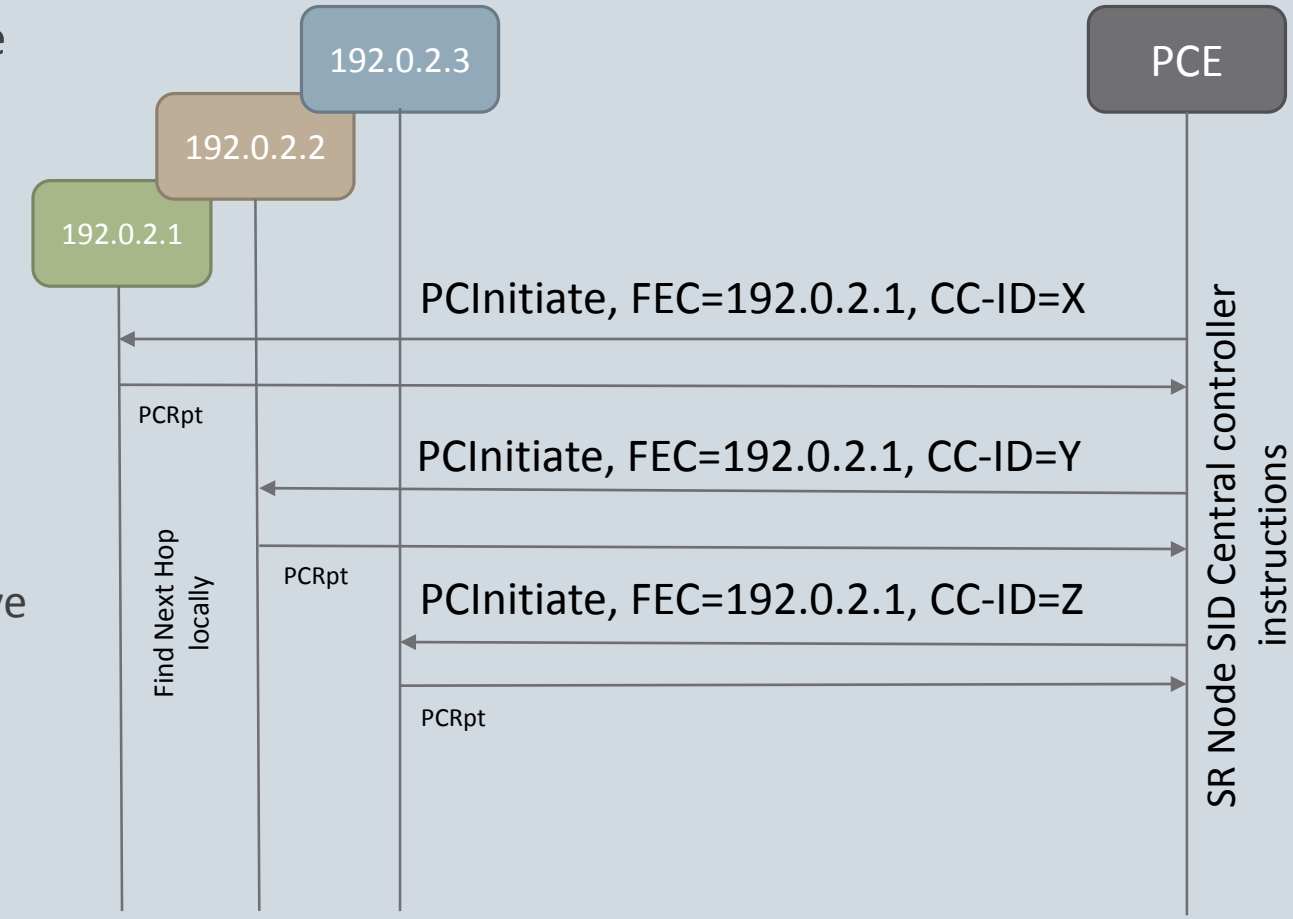
PCEP speaker can use any IP address while creating a TCP session.

- Important to link the TCP session IP address with the Router ID in TEDB for successful PCECC operations.



CCI: SR Node SID

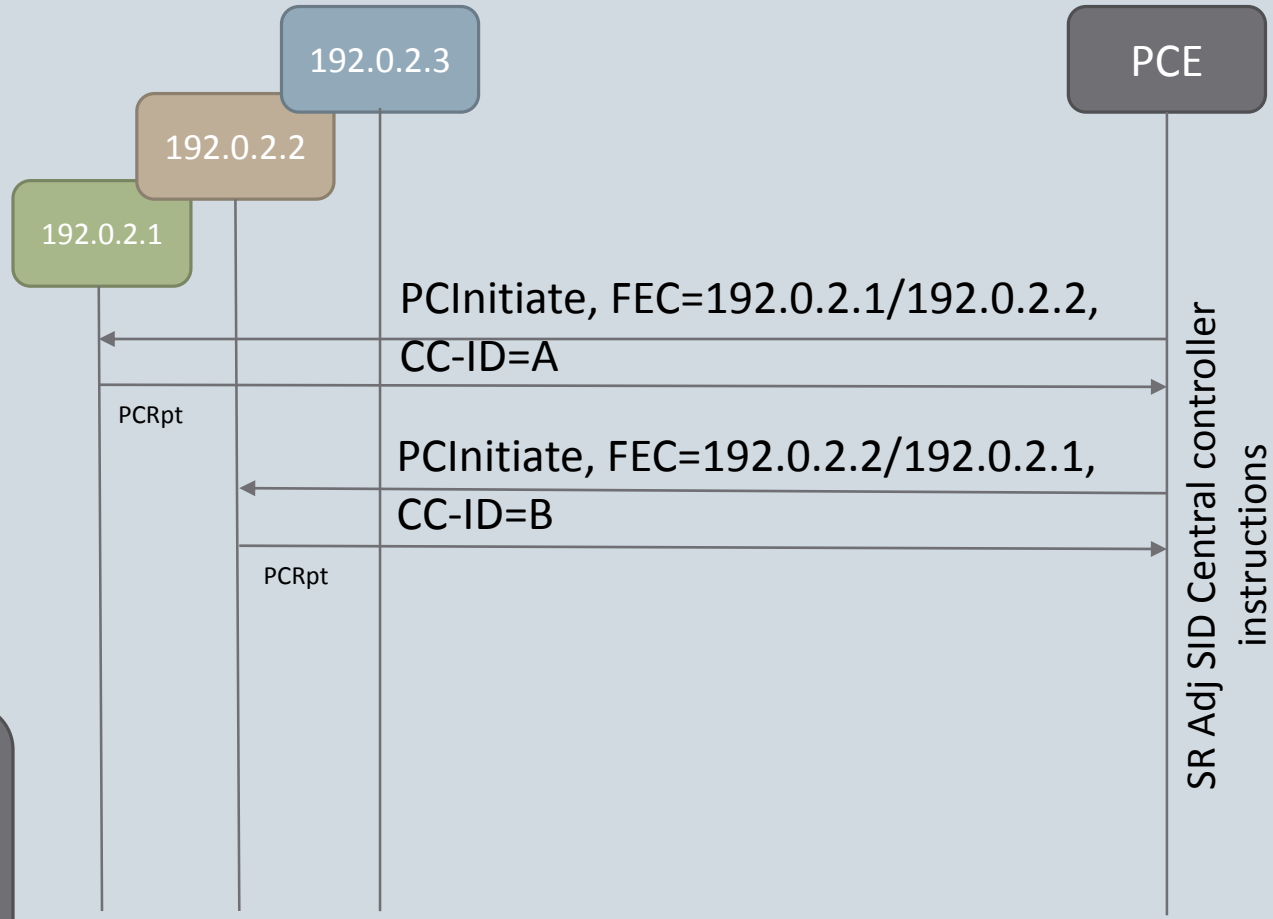
- Each node (PCC) is allocated a node-SID by the PCECC.
  - The PCECC sends PCInitiate message to update the label map of each node *(to all the nodes in the domain)*.
- PCC uses the local information (IGP) to determine the next-hop and download the label forwarding instructions accordingly.
  - It is the role of PCC to update the nexthop information as per the changes in network.
  - The PCInitiate message in this case does not have a LSP object but uses the new **FEC** object.
- In response to the PCInitiate, PCC sends the PCRpt message (which is also extended to include CCI)



CCI: SR Adj SID

- Each adjacency is allocated an Adj-SID by the PCECC.
  - The PCECC sends PCInitiate message to update the label map of the necessary nodes in the domain.
- The PCInitiate message in this case does not have a LSP object but uses the new **FEC** object.
- In response to the PCInitiate, PCC sends the PCRpt message (which is also extended to include CCI)

Once the node/adj SR SID are instructed to the nodes in form of CCI, the rest of the operation for SR in PCEP remains unchanged!



Re-delegation and Cleanup

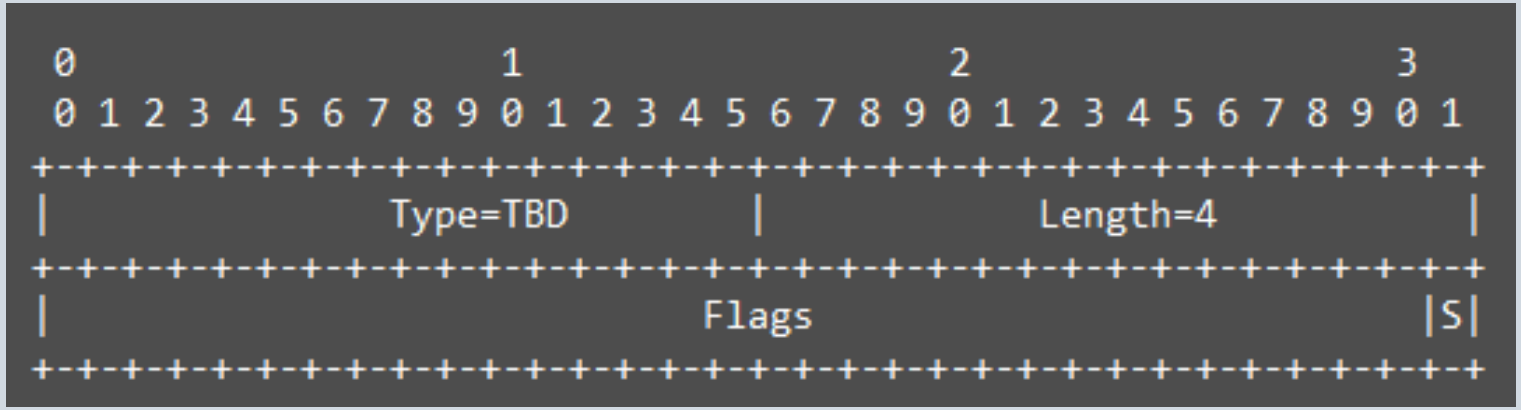
- When a new PCE takes over a PCECC LSP, it also gain control over the central controllers instructions in the same way by sending a PCInitiate message
  - that includes the SRP, LSP and CCI objects and carries the CC-ID and PLSP-ID identifying the instruction, it wants to take control of.
- The State Timeout Interval timer ensures that a PCE crash does not result in automatic and immediate disruption for the services.
  - Similarly the central controller instructions are not removed immediately upon PCE failure.
  - Instead, they are cleaned up on the expiration of this timer.
  - This allows for network cleanup without manual intervention.
  - The PCC support removal of CCI as one of the behaviors applied on expiration of the State Timeout Interval timer.

Synchronization of CCIs

- The purpose of Central Controllers Instructions synchronization (labels in the context of this document) is to make sure that the PCE's view of CCI (Labels) matches with the PCC's Label allocation.
  - This synchronization is performed as part of the LSP state synchronization.
- A PCC reports the state of its LSPs to the PCE using PCRpt messages after PCEP session establishment, and the PCE would initiate any missing LSPs and/or remove any LSPs that are not wanted.
- The same PCEP messages and procedure is also used for the Central Controllers Instructions synchronization.
  - The PCRpt message includes the CCI and the LSP object to report the label forwarding instructions.
  - The PCE would further remove any unwanted instructions or initiate any missing instructions.
- [I-D.litkowski-pce-state-sync] describes synchronization mechanism between the stateful PCEs.
  - The SR SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC SR state synchronization.
  - All PCEs MUST have a common view of all SR SIDs allocated in the domain.

Capability

- A new PST type for PCECC
- A new PCECC Capability sub-TLV is defined.
  - A S-bit is for PCECC-SR operations
- Stateful capability MUST also be exchanged for PCECC



PCEP Message

```
   <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>)|
                                           (<FEC>
                                           <CCI>)


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

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

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

Basic

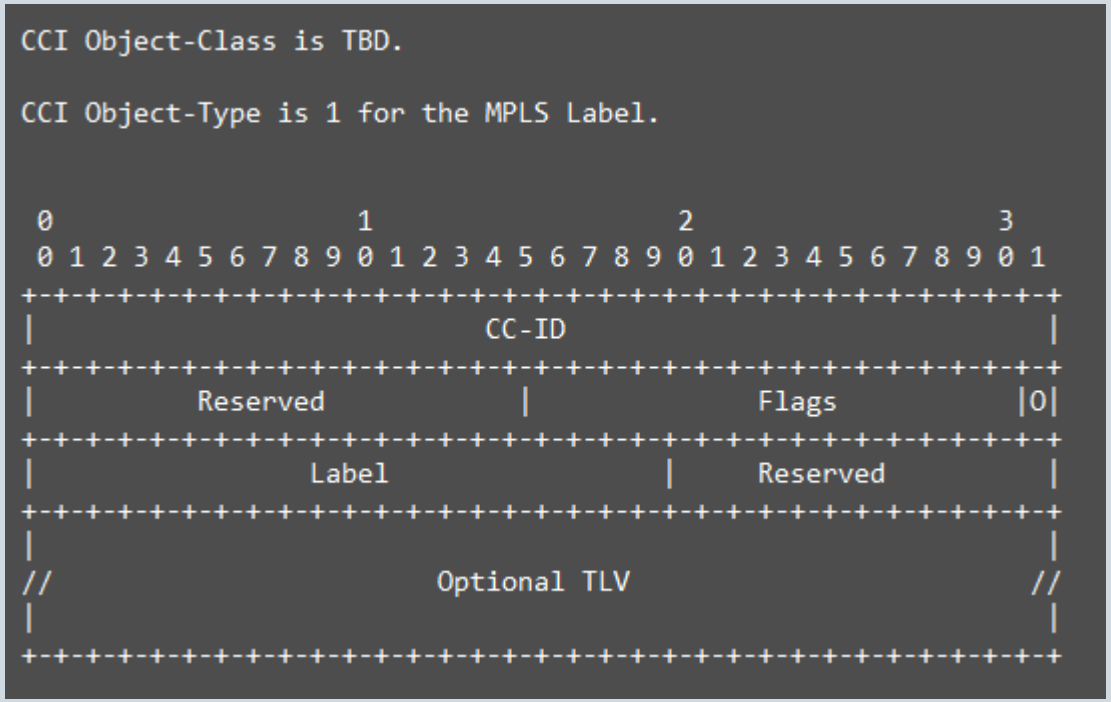
SR

```
   <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>)|
                                (<FEC>
                                <CCI>)


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


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

CCI Object

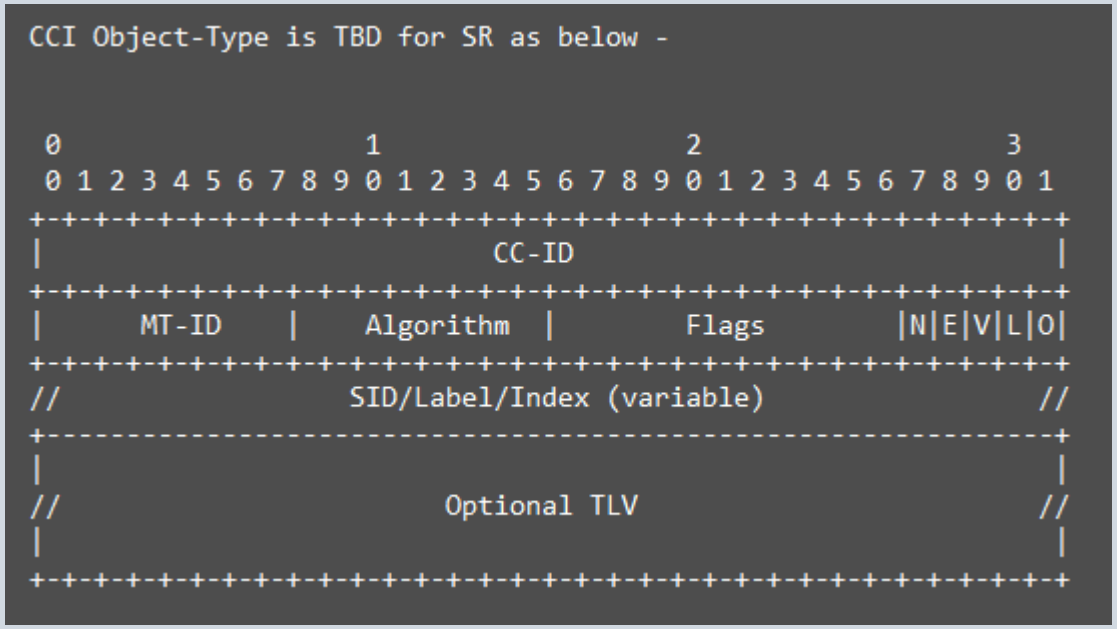
CCI Object-Class is TBD.

CCI Object-Type is 1 for the MPLS Label.



- O bit for out-label
- Address TLV for the next-hop information or local interface

CCI Object-Type is TBD for SR as below -

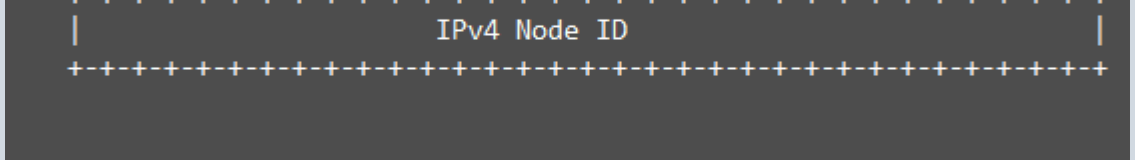


- Similar to SR fields
- L for Local/Global; V for Value/Index; E for explicit-null; N for No-PHP

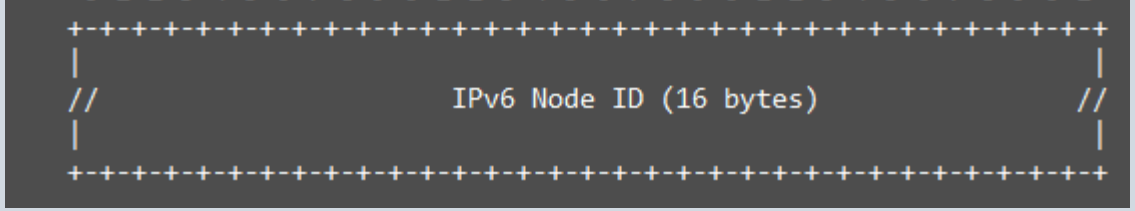
FEC Object

FEC Object-Class is TBD.

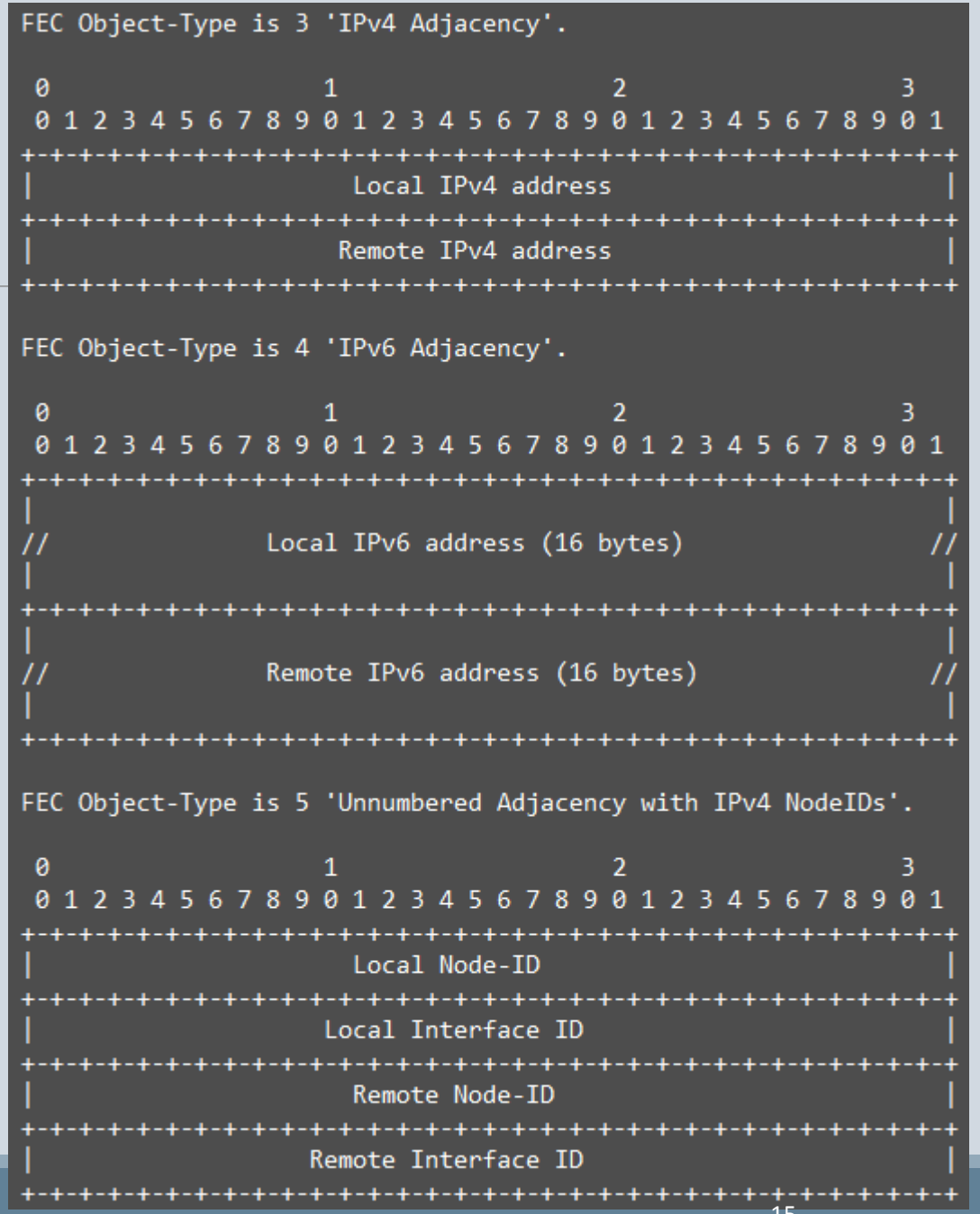
FEC Object-Type is 1 'IPv4 Node ID'.



FEC Object-Type is 2 'IPv6 Node ID'.



FEC Object-Type is 3 'IPv4 Adjacency'.



FEC Object-Type is 4 'IPv6 Adjacency'.

FEC Object-Type is 5 'Unnumbered Adjacency with IPv4 NodeIDs'.

Next Step

- Drafts have been refined to use the existing PCInitiate messages
  - instead of new messages defined earlier!
- Implementations of older version of draft exist
  - showcased in Hackathon and Bits-n-Bytes in past IETF!
- Aligned to RFC 8283 and other PCEP RFCs and drafts!
- Ready for WG adoption!


thank
you