        Stateful PCE Initiated LSP Setup Using Association Groups
              draft-shen-pce-lsp-setup-association-00

Abstract

   This document describes a hierarchical model for stateful PCE-
   initiated LSPs, based on the PCE association group mechanism. It
   defines two new association types, and introduces a generic mechanism
   for stateful PCE to use P2P LSPs as component LSPs (i.e. sub-LSPs) to
   construct and set up TE LSPs with complex path topology, including
   P2MP, MP2P, and MP2MP TE LSPs.

Status of This Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at http://datatracker.ietf.org/drafts/current/.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time. It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."

   This Internet-Draft will expire on April 11, 2015.

Copyright Notice

Table of Contents

1. Introduction

[RFC 5440] describes the Path Computation Element Protocol (PCEP) for
computation of Traffic Engineering Label Switched Path (TE LSP).
Stateful PCE [ietf-pce-stateful-pce] specifies a set of extensions to
PCEP to enable stateful control of TE LSPs for PCE. PCE-initiated
LSP [ietf-pce-pce-initiated-lsp] specifies the operation model for TE
LSPs that are initiated from PCE.

PCE association group [draft-minei-pce-association-group] further
introduces a generic mechanism for creating a grouping of LSPs. This
grouping can then be used to define associations between sets of LSPs
or between a set of LSPs and a set of attributes.

This document defines two new association types, namely "downstream
replication" and "downstream merge", for the PCE association group
mechanism. It introduces a generic mechanism for using PCE-initiate
P2P LSPs as component LSPs (i.e. sub-LSPs) to construct and set up
PCE-initiated LSPs with more complex path topologies, including P2MP,

MP2P, and MP2MP TE LSPs. This mechanism is applicable to PCE-
initiated TE LSPs. It adds a hierarchical model for PCE-initiated TE
LSPs.

2. Motivation

The motivation of this document is to introduce a hierarchical model
for stateful PCE-initiated LSPs, based on the PCE-initiated LSP
[ietf-pce-pce-initiated-lsp] and the PCE association group mechanism
[draft-minei-pce-association-group]. The mechanism uses P2P LSPs as
component LSPs (i.e. sub-LSPs) to set up TE LSPs with more complex
path topology, including P2MP, MP2P, and MP2MP TE LSPs.

3. Theory of operation

PCE association group [draft-minei-pce-association-group] describes a
generic mechanism for creating a grouping of LSPs. Base on this
mechanism, this document introduces two new association types:

   - Downstream replication

   - Downstream merge

With the new associations, a stateful PCE can use uni-directional and
bi-directional P2P LSPs as component LSPs (i.e. sub-LSPs) to
construct TE LSPs with complex path topology, including P2MP, MP2P,
and full-mesh and partial-mesh MP2MP TE LSPs. Specifically, the PCE
can first set up P2P sub-LSPs, and then apply the associations to
create desired forwarding state for the complex LSPs. In other
words, these complex LSPs are established as a group of associated
P2P sub-LSPs in a hierarchical manner. These LSPs and their sub-LSPs
are both PCE-initiated LSPs.

Each association group will create certain forwarding state on a PCC,
e.g. a set of forwarding entries, each with a set of nexthops. The
forwarding state remains stable during the lifetime of the
association group. In any case where there is change to the
relationship between member LSPs which will cause the forwarding
state to change, the PCE SHOULD create a new association group to
replace the old association group.

Note that all the sub-LSPs in this document are independent P2P LSPs.
They are not required to correlate with each other via RSVP
LSP_TUNNEL_IPv4/6 SESSION, LSP_TUNNEL_IPv4/6 SENDER_TEMPLATE,
P2MP_LSP_TUNNEL_IPv4/6 SESSION, or P2MP_LSP_TUNNEL_IPv4/6
SENDER_TEMPLATE object [RFC 3209, RFC 4857]. Their relationships are
formed by association groups at PCE level.

This section describes the nodal behaviors of these new associations, when they are applied to a group of LSPs on a node. Section 5 will describe a framework in a network scope how stateful PCE can use these associations while coordinating multiple nodes to set up complex LSPs.

Currently PCE association group [draft-minei-pce-association-group] only supports association groups for ingress LSPs. This document extends it to support association groups for transit and egress LSPs as well.

3.1. Downstream replication association

A downstream replication association group may include a number of ingress, transit and egress P2P LSPs. It allows a router (i.e. PCC) to create forwarding state with a so-called "flood nexthop", so that outgoing packets are replicated to and label-switched over every ingress and transit LSPs of the group. Note that not all combinations of ingress, transit and egress P2P LSPs are meaningful for forwarding. This document only considers the following specific combinations.

3.1.1. Group of ingress LSPs

In this scenario, the group includes purely ingress LSPs. The following example shows two P2P LSPs, A-B-C and A-D-E, and the label allocation schema at node A. Both LSPs are ingress LSPs at node A.

```
          ----B-----C
         /
        A
         \
          ----D-----E
```

```
      LSP        in-label      out-label
      --------------------------------
      A-B-C         -             100
      A-D-E         -             200
```
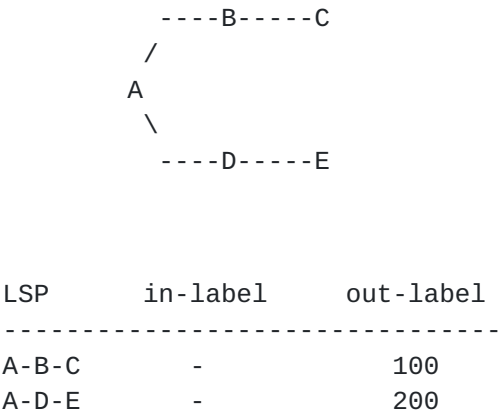
Figure 1

When downstream replication association is applied to the LSPs at node A, node A should create the following forwarding entry with a

flood nexthop. Thus, outgoing packets will be replicated and label-
switched along both LSPs.

```
              in-label     out-label    out-interface
              ---------------------------------------
                  -            100            to-B
                               200            to-D
```

                          Figure 2

3.1.2. Group of transit and ingress LSPs

In this scenario, the group may include [1, n) transit LSPs and [0,
n) ingress LSPs, but no egress LSP. One transit LSP MUST be
designated to have its incoming packets replicated to every other
LSPs of the group, in addition to being label-switched along the
transit LSP itself. If there are multiple transit LSPs in the group,
the designated LSP MUST be marked by a "Designated in-LSP" flag of
the Type-Specific Flags in Association object, defined in Section 4.
Incoming packets on other transit LSPs are discarded.

The following example shows three P2P LSPs, A-B-C-D, E-B-F-G, and
B-X-Y, and the label allocation schema at node B. At node B, LSP
A-B-C-D and E-B-F-G are transit LSPs, and LSP B-X-Y is a ingress LSP.

```
                         ---F-----G
                        /
                       /
                A-----B-----C-----D
                     / \
                    /   \
                 E---     ---X-----Y
```
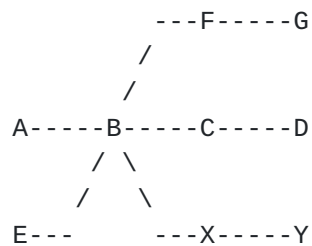
```
           LSP        in-label     out-label
           -------------------------------
           A-B-C-D     100           200
           E-B-F-G     300           400
           B-X-Y        -            500
```

                          Figure 3

When downstream replication association is applied to the LSPs at
node B, with LSP A-B-C-D specified as the designated in-LSP, node B
should create the following forwarding entry for label 100 with a
flood nexthop, and discard entry for label 300. Thus, packets
received on LSP A-B-C-D will be replicated and label-switched along
all LSPs.


```
          in-label      out-label   out-interface
          ----------------------------------------
          100             200            to-C
                          400            to-F
                          500            to-X
          300                __discard__
```


Figure 4

3.1.3. Group of a single egress LSP and other LSPs

In this scenario, the group includes a single egress LSP, [0, n)
transit LSPs, and [0, n) ingress LSPs. The egress LSP must be UHP
(ultimate hop popping). Incoming packets on the egress LSP are
replicated to and label-switched along every other LSPs of the group.
Incoming packets on the transit LSPs are discarded.

The following example shows four P2P LSPs, A-B-C, C-D-E, F-G-C-H-I,
and C-X-Y, and the label allocation schema at node C. LSP A-B-C is a
UHP LSP, and an egress LSP at node C. LSP C-D-E and C-X-Y are both
ingress LSPs at node C. LSP F-G-C-H-I is a transit LSP at node C.

```
                        ---H-----I
                       /
                      /
             A-----B-----C-----D-----E
                     / \
                    /   \
             F-----G---     ---X-----Y



             LSP       in-label     out-label
             --------------------------------
             A-B-C       100            -
             C-D-E        -            200
             C-X-Y        -            300
             F-G-C-H-I  400            500
```
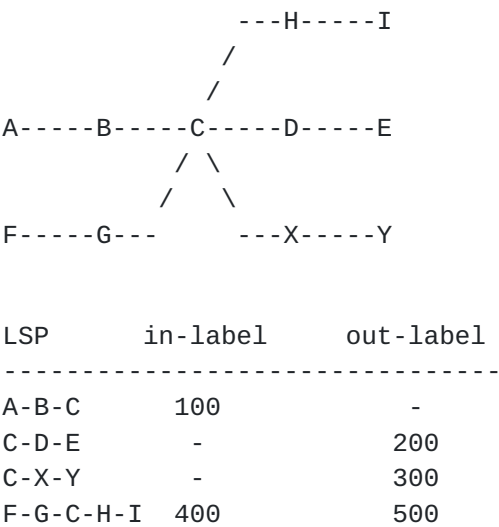
                         Figure 5

   When downstream replication association is applied to the LSPs at
   node C, node C should create the following forwarding entry for label
   100 with a flood nexthop, and discard entry for label 400. Thus,
   packets received on LSP A-B-C will be replicated and label-switched
   along the other LSPs.

```
             in-label     out-label    out-interface
             ---------------------------------------
             100            200             to-D
                            300             to-X
                            500             to-H
             400              __discard__
```

                         Figure 6

3.2. Downstream merge association

   An downstream merge association group includes a single transit P2P
   LSP, [1, n) egress P2P LSPs, and no ingress LSP. All the egress LSPs
   must be UHP. This type of association group allows a router (i.e.
   PCC) to set up forwarding state in such a fashion that packets
   received on any LSP of the group will be label-switched along that
   transit LSP. In other words, multiple incoming flows from upstream
   are merged into a single flow downstream.

The following example shows two P2P LSPs, A-B-C-D and X-Y-B, and the label allocation schema at node B. Note that LSP X-Y-B is a UHP LSP in this case.


```
        A-----B-----C-----D
              |
              |
        X-----Y
```


```
    LSP       in-label     out-label
    -------------------------------
    A-B-C-D    100            200
    B-X-Y      300             -
```
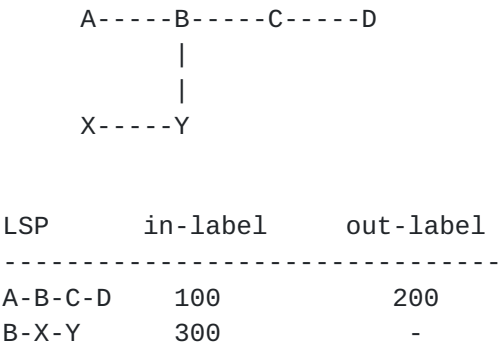

                        Figure 7

When downstream merge association is applied the LSPs at node B, B should then create the following forwarding entries. Thus, incoming packets on both LSPs will be label-switched downstream along LSP A-B-C-D.


```
        in-label     out-label   out-interface
        ----------------------------------------
        100           200           to-C
        300           200           to-C
```


                        Figure 8

4. PCEP extensions

PCE association group [draft-minei-pce-association-group] defines the Association object. This document defines two new Association types for the Association object. The values of these types are TBD.

   - Downstream Replication

   - Downstream Merge

For Downstream Replication type, this document also allocates a flag, i.e. the Designated in-LSP flag in the Type-specific Flags field. When there are more than one transit LSPs in a group (Section 3.1.2),

this flag MUST be set for the designated transit LSP, whose incoming
packets should be replicated to every transit and ingress LSPs of the
group. There SHOULD be one and only one transit LSP with this flag
set.

Per [draft-minei-pce-association-group], the format of the
Association object for Downstream Replication and Downstream Merge is
shown Figure-9:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Type   |  Generic flags   |R| Type-specific flags          |D|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Association group id                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//            Optional TLVs                                   //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
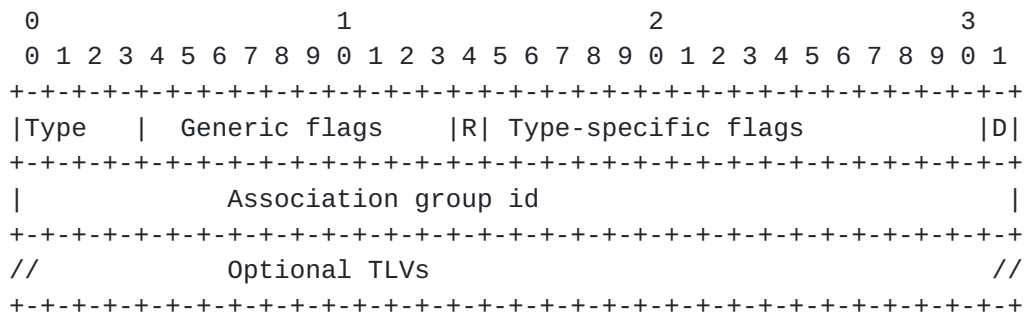
Figure 9

Type - Downstream Replication or Downstream Merge. Values TBD.

Generic flags - flags for the association object. The R flag
indicating removal from the association group.

Type-specific flags - specific to the association type. The D
flag is defined for Designated in-LSP.

Association group id - identifier of the association group.

Error handling for these new association types will be defined in a
future version of this document.

5. PCE-initiated LSPs in a hierarchical model

This section describes a framework for stateful PCE to use the
downstream replication and downstream merge associations to construct
and set up TE LSPs with complex path topology in a hierarchical
manner.

5.1. Branch-node initiated P2MP TE LSPs

RSVP P2MP TE LSP is specified by RFC 4857, where a P2MP LSP is
established by its source node signaling an S2L (source-to-leaf) sub-
LSP to each leaf node. This signaling mechanism can be viewed as a
"source-node initiated" P2MP LSP model. Stateful PCE for P2MP LSPs
[draft-palle-pce-stateful-pce-p2mp-04] introduces PCEP extensions to
support stateful PCE for this model.

One observation about the source-node initiated P2MP LSP model is
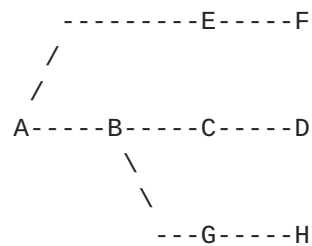that the source node must be provisioned with all leaf nodes and
signal S2L sub-LSPs to all the leaf nodes, regardless of the actual
number of traffic flows branching out from it. Similarly, a transit
router must participate in the signaling of all the S2L sub-LSPs
traversing it, regardless of the actual number of traffic flows
branching out from it. This leaves some room of improvement for
efficiency.

With stateful PCE, an alternative P2MP LSP model becomes available.
It is called "branch-node initiated" P2MP LSP model. In this model,
we introduce the notion of "branch-node to leaf" (B2L) sub-LSP, which
represents a single-path segment of a P2MP LSP that starts from a
branch node and ends at a leaf node. The branch node is the headend
of this B2L sub-LSP, while the leaf node is the tailend.

Therefore, a P2MP LSP can be viewed as to consist of a set of S2L
sub-LSPs (called level-1 sub-LSPs) whose ingress is the source node,
a set of B2L sub-LSPs (i.e. level-2 sub-LSPs) attached to the level-1
sub-LSPs at some branch nodes, a set of B2L sub-LSPs (i.e. level-3
sub-LSPs) attached to the level-2 sub-LSPs at some branch nodes, and
so on. A branch node is called a level-N branch node (N > 1), if
some level-N sub-LSPs are attached to a level-(N-1) sub-LSP at this
node. The branch node is the ingress node of the level-N sub-LSPs,
and a transit node of the level-(N-1) sub-LSP. For example:

```
                    --------E-----F
                  /
                 /
               A-----B-----C-----D
                     \
                      \
                       ---G-----H


               Source node: A
               Leaf nodes: D, F, H
               Branch node: B
               Level-1 sub-LSPs: A-B-C-D, A-E-F
               Level-2 sub-LSPs: B-G-H
```

                              Figure 10

   With this model, a stateful PCE can construct and set up a P2MP LSP
   by two steps. First, initiate S2L and B2L sub-LSP creation on the
   source node and branch nodes, respectively. Second, apply downstream
   replication association to the sub-LSPs to create desired flood
   nexthops on those nodes. Note that the sub-LSPs are not required to
   use P2MP_LSP_TUNNEL_IPv4/6 SESSION or P2MP_LSP_TUNNEL_IPv4/6
   SENDER_TEMPLATE [RFC 4857] to correlate with each other. Instead,
   they are signaled as independent P2P LSPs. It is the downstream
   replication association that groups them together explicitly at PCEP
   level.

   The partitioning of a P2MP LSP into sub-LSPs should be performed by
   the stateful PCE as part of P2MP path computation. Hence, it is
   outside the scope of this document. Note that for a given P2MP LSP,
   there may be multiple schemas to partition it into sub-LSPs. After
   the PCE has computed the sub-LSP paths, it sends PCinit messages to
   the source node and branch nodes to for the sub-LSPs. The source
   node and branch nodes then signal these sub-LSPs independently.
   After the sub-LSPs are set up as indicated by PCrpt messages, the PCE
   sends PCupd messages to the source node and branch nodes to create
   downstream replication association groups. The PCupd messages MUST
   carry an Association object with the new Downstream Replication
   association type defined in this document. The Association Group ID
   should be non-0 and unique within each node. Based on this
   association, the source node MUST build a flood next-hop by combining
   the downstream labels and interfaces of all level-1 sub-LSPs;
   Likewise, each level-N branch node MUST build a flood next-hop for
   the in-label of the level-(N-1) sub-LSP, by combining the downstream
   labels and interfaces of the level-(N-1) and all level-N sub-LSPs.

In the example above, assume the following label allocation schema at
the source node A and branch node B.


```
Sub-LSP   node   in-label      out-label
-----------------------------------------
A-B-C-D    A        -            100
A-E-F      A        -            200
A-B-C-D    B       300           400
B-G-H      B        -            500
```

Figure 11

After downstream replication association is applied to the sub-LSP on
node A and B, node A and B will create the following forwarding
entries with flood nexthop, respectively.


```
Node A:

Group ID: 1
Members: A-B-C-D, A-E-F

in-label     out-label   out-interface
---------------------------------------
   -            100           to-B
                200           to-E

Node B:

Group ID: 1
Members: A-B-C-D, B-G-H

in-label     out-label   out-interface
---------------------------------------
  300           400           to-C
                500           to-G
```

Figure 12

As shown above, the number of sub-LSPs that a node needs to be aware
of is the number of traffic flows branching out from it, rather than
the total number of leaf nodes downstream of it. This means less

   sub-LSPs to maintain in RSVP. Hence, the overall efficiency and
   scalability of signaling is improved.

5.2. MP2P TE LSPs

   MP2P RSVP-TE [draft-yasukawa-mpls-mp2p-rsvpte] specifies a mechanism
   and a set of RSVP extensions for signaling MP2P LSPs from multiple
   ingress nodes to a single egress node. With stateful PCE and PCE
   association group, an alternative MP2P LSP model becomes possible,
   without the need for RSVP extension. In this model, we call a sub-
   LSP that goes from an ingress node to the egress node an I2E (i.e.
   ingress-to-egress) sub-LSP. We also introduce the notion of "ingress
   to merge-node" (I2M) sub-LSP. It refers to a single-path segment of
   an MP2P LSP that starts from an ingress node and ends at a merge
   node. The ingress node is the headend of this I2M sub-LSP, while the
   merge node is the tailend.

   Therefore, an MP2P LSP can be viewed as to consist of a set of I2E
   sub-LSPs (called level-1 sub-LSPs), a set of I2M sub-LSPs (i.e.
   level-2 sub-LSPs) merged with the I2E sub-LSPs, a set of I2M sub-LSPs
   (i.e. level-3 sub-LSPs) merged with the level-2 sub-LSPs, and so on.
   A merge node is called a level-N merge node (where N > 1), if one or
   multiple level-N sub-LSPs merge with a level-(N-1) sub-LSP at this
   node. In this case, the merge node is the egress node of the level-N
   sub-LSP(s), while it is a transit node of the level-(N-1) sub-LSP.
   For example:

```
                  A------B------C------D
                                |
                                |
                  E------F------
                         |
                         |
                  G------H


                  Ingress nodes: A, E, G
                  Egress node: D
                  Merge nodes: C, F
                  Level-1 sub-LSP: A-B-C-D
                  Level-2 sub-LSP: E-F-C
                  Level-3 sub-LSP: G-H-F
```
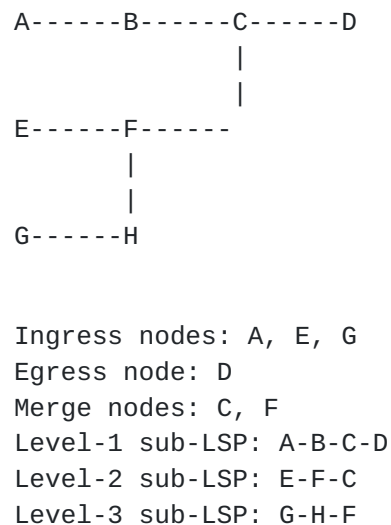
                          Figure 13

With this model, a stateful PCE can set up an MP2P LSP by two steps.
First, initiate sub-LSP creation at every ingress node. All ingress
nodes signal I2E and I2M sub-LSPs as independent P2P LSPs. All I2M
sub-LSPs are signaled as UHP LSPs. Second, apply downstream merge
association to the sub-LSPs at every merge node. Note that at each
level-N merge node, there should be one level-(N-1) transit sub-LSP
and one or multiple level-N egress sub-LSP. This merge node MUST
create normal forwarding state for the level-(N-1) transit sub-LSP.
For each level-N egress sub-LSP, the merge node MUST create
forwarding state in such a manner that the incoming packets of this
sub-LSP are also label-switched downstream along the level-(N-1)
transit sub-LSP.

The partitioning of an MP2P LSP into sub-LSPs should be performed by
the stateful PCE as part of MP2P path computation. Hence, it is
outside the scope of this document. Note that for a given MP2P LSP,
there may be multiple schemas to partition it into sub-LSPs. After
the PCE has computed the sub-LSP paths, it sends PCinit messages to
the ingress nodes. The ingress nodes then signal these sub-LSPs
independently. All I2M sub-LSPs must be signaled as UHP LSPs. After
the sub-LSPs are set up as indicated by PCrpt messages, the PCE sends
PCupd messages to merge nodes to create downstream merge association
groups. The PCupd message should carry an Association object with
the new Downstream Merge association type defined in this document.
The Association Group ID should be non-0 and unique within each merge
node. Each level-N merge node MUST create forwarding state for each
sub-LSP of the group, using a common nexthop with the downstream
label and interface of the level-(N-1) transit sub-LSP.

In the above example, assume the following schema of label allocation
at merge nodes C and F.


```
          Sub-LSP   node   in-label     out-label
          -------------------------------------------
          A-B-C-D    C        100      implicit null
          E-F-C      C        200           -
          E-F-C      F        300          200
          G-H-F      F        400           -
```


                          Figure 14

After downstream merge association is applied to the sub-LSPs on node
C and F, node C and F will create the following forwarding entries,
respectively.

```
            Node C:

            Group ID: 1
            Members: A-B-C-D, E-F-C

            in-label     out-label   out-interface
            ----------------------------------------
               100          pop             to-D
               200          pop             to-D

            Node F:

            Group ID: 1
            Members: E-F-C, G-H-F

            in-label     out-label   out-interface
            ----------------------------------------
               300          200             to-C
               400          200             to-C
```

                         Figure 15

5.3. MP2MP TE LSPs

   RSVP-TE currently doesn't provide a mechanism for signaling MP2MP TE
   LSPs. An MP2MP LSP may be constructed by setting up separate P2MP
   LSPs from ingress nodes. However, these P2MP LSPs do not have the
   awareness of each other. Hence they are unable to share network
   resources on common links or nodes. Also, the number of sub-LSPs
   needed for an MP2MP LSP is generally bound to the average number of
   ingress nodes multiplied by the average number of egress nodes. In
   the case of a full-mesh MP2MP LSP with N ingress/egress nodes, this
   number is $(N-1)^2$.

   With the downstream replication and downstream merge associations, a
   stateful PCE can construct and set up full-mesh and partial-mesh
   MP2MP LSPs by using uni-directional and bi-directional P2P LSPs as
   component LSPs. The number of P2P LSPs needed can be significantly
   lower than that of the above model using P2MP LSPs. In some network
   topologies, O(n) may be achievable.

   The idea of constructing an MP2MP LSP is similar to that of P2MP and
   MP2P LSPs. A stateful PCE computes an MP2MP path, which is normally
   a graph, consisting of a set of nodes and unidirectional and
   bidirectional edges. The PCE then partitions the graph into a set of
   unidirectional and bidirectional single-path segments, branching out,
   intersecting or merging with each other at some nodes. These nodes

are generally called stitching nodes in this document. The algorithm
for performing such partitioning and identifying the stitching nodes
is outside the scope of this document. The stateful PCE then models
these segments as unidirectional and bidirectional P2P sub-LSPs.
Note that bidirectional P2P sub-LSPs are not mandatory, because they
can be achieved by using co-routed unidirectional P2P LSPs.

The PCE then sends PCinit messages to the ingress nodes and stitching
nodes to initiate the creation of these sub-LSPs. The ingress nodes
and stitching nodes signal the sub-LSPs independently. All sub-LSPs
that egress at an stitching node must be signaled as UHP LSPs. After
the sub-LSPs are set up as indicated by PCrpt messages, the PCE sends
PCupd messages to the stitching nodes to create downstream
replication and downstream merge association groups. The PCupd
messages should carry one or multiple Association objects with the
Downstream Replication or Downstream Merge association type defined
in this document. Each stitching node then creates forwarding state
based on the associations, to achieve the desired forwarding state
for the MP2MP LSP.

Consider an example where a stateful PCE has computed the following
path for a full-mesh MP2MP LSP with three ingress/egress nodes, A, E,
and F. The stateful PCE then partitions the path into four
unidirectional P2P sub-LSPs, i.e. A-B-C-D-E, E-D-C-B-A, F-G-C, and
C-G-F. Alternative, the stateful PCE may partition the path into two
bi-directional P2P sub-LSPs, i.e. A-B-C-D-E and C-G-F. In either
case, node C is a stitching node, and label allocation for the sub-
LSPs can remain the same.

```
                    A-----B-----C-----D-----E
                                |
                                |
                                G
                                |
                                |
                                F


            Ingress/egress nodes: A, E, F
            Sub-LSPs: A-B-C-D-E, E-D-C-B-A, F-G-C, C-G-F
            Stitching node: C
```

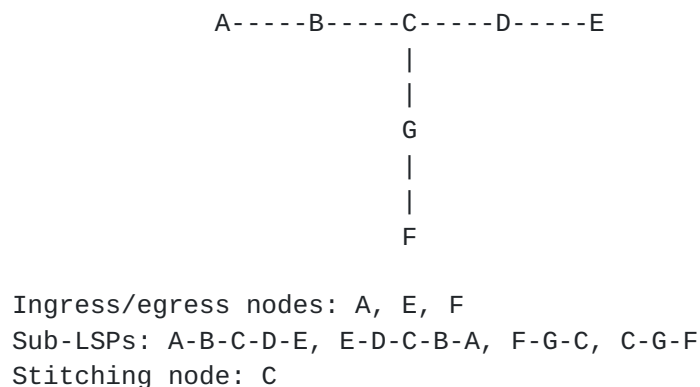   Figure 16

Assume the following label allocation schema at node C, after the
sub-LSPs are set up. In particular, the sub-LSP F-G-C is a UHP LSP.

```
            Sub-LSP          in-label     out-label
            ----------------------------------------
            A-B-C-D-E           101          102
            E-D-C-B-A           201          202
            F-G-C               301           -
            C-G-F                -           402
```

                          Figure 17

   The stateful PCE applies the following downstream replication
   association groups to the sub-LSPs at node C.

                  Group ID: 1
                  Members: A-B-C-D-E, C-G-F

                  Group ID: 2
                  Members: E-D-C-B-A, C-G-F

                  Group ID: 3
                  Members: A-B-C-D-E, E-D-C-B-A

                          Figure 18

   Base on the association groups, node C creates the following
   forwarding entries with flood nexthop.


```
         In-label      out-label       out-interface
         --------------------------------------------
         101             102               to-D
                         402               to-G
         --------------------------------------------
         201             202               to-B
                         402               to-G
         --------------------------------------------
         301             102               to-D
                         202               to-B
```

                          Figure 19

5.4. Other considerations

   It is expected that not all routers in a network can act as PCC or
   support the mechanism in this document. A PCE SHOULD use the set of
   routers that support this mechanism as a constraint in path
   computation, and only use these routers as branch, merge, and
   stitching nodes.

   When there is a change to the path topology of an LSP, a PCE may
   perform an incremental modification on the LSP, by adding a set of
   new sub-LSPs and association groups, and deleting a set of existing
   sub-LSPs and association groups that are no longer valid.
   Alternatively, the PCE may simply create new sub-LSPs and association
   groups for the entire LSP as a new LSP, and replace the existing LSP
   in a make-before-break fashion.

6. Failure protection

   Several strategies can be used provide link and node failure
   protection for LSPs set up by the mechanism in this document.

   The first strategy is based on the fast re-route (aka. local repair)
   specified in RFC 4090. This is driven by routers at PLR (point of
   local repair) using pre-installed backup forwarding state, and has
   the advantage of fast reaction. It is suitable for protection
   against link failures and non-branch/merge/stitching node failures.
   On an LSP set up by the mechanism in this document, each link is
   traversed in each direction by one and only one sub-LSP; Likewise,
   each non-branch/merge/stitching node is traversed in each direction
   by one and only one sub-LSP. Therefore, protecting the LSP against a
   link failure or non-branch/merge/stitching node failure is no
   different than protecting a single sub-LSP against that failure.

   The second strategy is global repair driven by PCE. A PCE computes
   an alternative P2MP, MP2P or MP2MP path based on the new topology,
   and constructs a new LSP with new sub-LSPs and association groups to
   replace the existing LSP. This strategy is applicable to protection
   against branch/merge/stitching node failures, where the above fast
   re-reroute is not possible due to PLR's lack of knowledge of
   downstream sub-LSPs to build backup forwarding state. It can also be
   used for protection against link failure and non-branch/merge/
   stitching node failure, when fast re-reroute is not available.
   However, it is worth noting that a global repair may take time, due
   to path recomputation and LSP reestablishment.

7. OAM

    The operation of OAM including LSP ping will be described in a future
    version of this document.

8. IANA considerations

    This document requests IANA to allocate values for the Downstream
    Replication association type and Downstream Merge association type
    defined in this document.

9. Security Considerations

    The security consideration described in [ietf-pce-pce-initiated-lsp]
    and [draft-minei-pce-association-group] apply to this document.
    Additional consideration related to malicious PCE are introduced in
    this document, as the PCE may now modify forwarding state on the PCC
    through downstream replication association and downstream merge
    association.

10. Acknowledgements

    The author would like to thank David Wood and Colby Barth for their
    kind review and valuable comments.

11. References

11.1. Normative References

    [draft-minei-pce-association-group]
              Minei, I., Crabbe, E., Sivabalan, S., Zhang, X., and Y.
              Tanaka, "PCE extensions for establishing relationships
              between sets of LSPs", draft-minei-pce-association-group
              (work in progress), 2014.

    [ietf-pce-pce-initiated-lsp]
              Crabbe, E., Minei, I., Sivabalan, S., and R. Verga, "PCEP
              extensions for PCE-initiated LSP setup in a stateful PCE
              model", draft-ietf-pce-pce-initiated-lsp (work in
              progress), 2014.

    [ietf-pce-stateful-pce]
              Crabbe, E., Minei, I., Medved, J., and R. Verga, "PCEP
              extensions for stateful PCE", draft-ietf-pce-stateful-pce
              (work in progress), 2014.

    [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5226]   Narten, T. and H. Alvestrand, "Guidelines for Writing an
            IANA Considerations Section in RFCs", BCP 26, RFC 5226,
            May 2008.

[RFC5440]   Vasseur, JP. and JL. Le Roux, "Path Computation Element
            (PCE) Communication Protocol (PCEP)", RFC 5440, March
            2009.

[RFC4578]   Johnston, M. and S. Venaas, "Dynamic Host Configuration
            Protocol (DHCP) Options for the Intel Preboot eXecution
            Environment (PXE)", RFC 4578, November 2006.

[draft-yasukawa-mpls-mp2p-rsvpte]
            Yasukawa, S., "Supporting Multipoint-to-point LSPs in MPLS
            TE", draft-ietf-pce-stateful-pce (work in progress), 2010.

11.2. 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, December 2001.

[RFC4655]   Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
            Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC4657]   Ash, J. and J. Le Roux, "Path Computation Element (PCE)
            Communication Protocol Generic Requirements", RFC 4657,
            September 2006.

[RFC4090]   Pan, P., Swallow, G., and A. Atlas, "Fast Reroute
            Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May
            2005.

Author's Address

   Yimin Shen
   Juniper Networks
   10 Technology Park Drive
   Westford, MA 01886
   USA

   Phone: +1 9785890722
   Email: yshen@juniper.net