                    Advertising LSPs with Segment Routing
                 draft-bowers-spring-advertising-lsps-with-sr-00

Abstract

   Segment routing uses globally-known labels to accomplish forwarding
   along shortest paths, and label stacks to accomplish explicit routing
   along arbitrary paths. These labels are advertised using an IGP.
   This draft describes how label bindings corresponding to RSVP, LDP,
   BGP labeled-unicast, and static LSPs are advertised in segment
   routing and how these labels can be combined with other segment
   routing labels to create forwarding paths. This draft also describes
   how context labels for egress node protection are advertised in using
   segment routing IGP extensions.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   [I-D.filsfils-spring-segment-routing] describes the segment routing
   architecture. In segment routing, LDP-like forwarding behavior along
   shortest paths is achieved using globally-known node labels.
   Globally-known node labels can be distributed in one of two ways.
   Each router can directly advertise a globally unique node label in
   the IGP. Or each router can advertise a globally unique node index
   value as well a locally assigned label block, allowing any router in
   the IGP area to determine the mapping of locally-assigned label to
   globally unique node index for any other router in the area.

   In order to forward traffic along explicit paths, segment routing
   uses stacks of adjacency labels. Each router uses the IGP to
   advertise locally significant adjacency labels corresponding to each
   of the router's outgoing interfaces. This allows any router in the
   IGP area to construct an arbitrary forwarding path by imposing a
   stack of adjacency labels on a packet. Forwarding is accomplished at
   each router by reading the top label on the stack to determine next-
   hop interface (based on its adjacency label to interface mapping),

popping that top label, and forwarding the packet out the next-hop
interface.

The above is only a short description of the use of node and
adjacency labels in segment routing. See
[I-D.filsfils-spring-segment-routing] for more detail on node and
adjacency label semantics as well as combining node and adjacency
labels in a label stack.

In addition to node and adjacency label advertisements, which
advertise label bindings corresponding to nodes and interfaces, it is
useful to advertise more abstract label bindings in the IGP, using a
Binding advertisement. This draft describes how label bindings
corresponding to RSVP, LDP, BGP labeled-unicast, and static LSPs are
advertised in segment routing and how these labels can be combined
with other segment routing labels to create forwarding paths. This
draft also describes how context labels for egress node protection
are advertised using a Binding advertisement.

2. Segment routing label binding advertisements

An LSP and its associated label is advertised in the IGP using the
Binding advertisement extensions defined in
[I-D.ietf-isis-segment-routing-extensions] and
[I-D.ietf-ospf-segment-routing-extensions]. The router which is the
ingress for the LSP advertises the label as well as the Forwarding
Equivalency Class(FEC) associated with the LSP. The advertisement
may include other information that describes the LSP. An Explicit
Route Object(ERO) may be included to describe the path taken by the
LSP. An ERO metric value may be included to indicate the cumulative
IGP or TE metric associated with the LSP along the path described by
the ERO.

Consider the network shown in Figure 1. A router R4 advertising a
label L1 and a FEC P using a Binding advertisement is indicating the
following forwarding behavior: Assume that a packet arrives at R4
with a label stack of the form <L1,L2,...,LN>, where n>=1 and L1 is
the top label. R4 will remove L1 from the label stack and the packet
will be forwarded to the router associated with FEC P (call it R7).
The precise set of MPLS label operations that get the packet from R4
to R7 is not well-defined. However, the label stack when the packet
arrives at R7 is well-defined. If n=1, then the packet will arrive
at R7 with the underlying packet header exposed, while if n>1, then
the packet will arrive at R7 with the label stack <L2,...,LN>,. (R7
may advertise an explicit null label in the underlying MPLS label
distribution, in which case R7 will know to remove the explicit null
label.)

```
          +-----+         +-----+         +-----+
          |     |         |     |         |     |
      R1--|     |--R4--|     |--R7--|     |--R10
          |     |         |     |         |     |
          +-----+         +-----+         +-----+
```
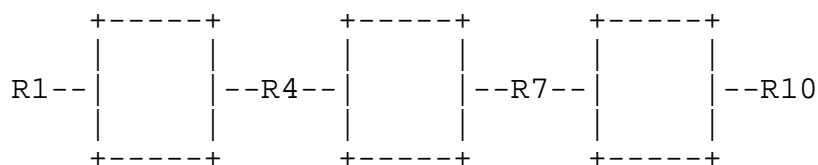
Figure 1: Example network

Continuing with the example above, a segment routing capable router
R1 uses the information describing the LSP contained in the Binding
advertisement (FEC, ERO, metric, etc.) to determine if it wants to
use that LSP as part of a longer forwarding path. If so, R1 uses the
label L1 advertised by R4 for the LSP in the construction of a label
stack. R1 can use any combination of segment routing labels stacked
above L1 to define a path to reach R1. The only requirement is that
label L1 be exposed as the top label when the packet reaches R4.
This will ensure that R4 forwards the packet to R7. If R1 has
determined that R7 supports segment routing, then R1 can also include
additional segment routing labels below L1 in the label stack to
specify the forwarding path beyond R7.

The description above assumes that R1 is responsible for computing
the forwarding path and the associated label stack. However, the
same forwarding behavior can be achieved if a centralized controller
is used to compute the path and communicate the associated label
stack to R1 via PCEP with the appropriate segment routing extensions
(see [I-D.sivabalan-pce-segment-routing]). The same is true of the
examples for specific label distribution protocols provided below.

3. Conventions used in the following examples

To simplify the diagrams and descriptions of the examples in this
draft, we assume that all routers advertise a router-id of the form
10.0.0.XX, where XX is the router number of the form RXX. For
example, in Figure 2 R16 advertises router-id=10.0.0.16 as a loopback
address.

Unless otherwise stated, links between routers all have the same IGP
metric of 10.

The node-SID index value for a router with name RXX will be XX. For
example, R23 in Figure 2 has an index value of 23. We assume that
all routers are advertising the same SR Global Block of 1000-1099.
For example, for R11 Figure 2 to send a packet to R23 on the shortest
path, R11 sends the packet to R13 with the top label equal to 1023.

4. Advertising an RSVP-TE LSP

```
                       +-R35-+
                      /       \
         R11-------R13--R14--R15--R16--R17--R18
          |         |    |    |    |    |    |
         R21--R22--R23--R24--R25--R26--R27--R28

         |<--- SR --->|              |<--- SR --->|
                 |<------RSVP------>|
```
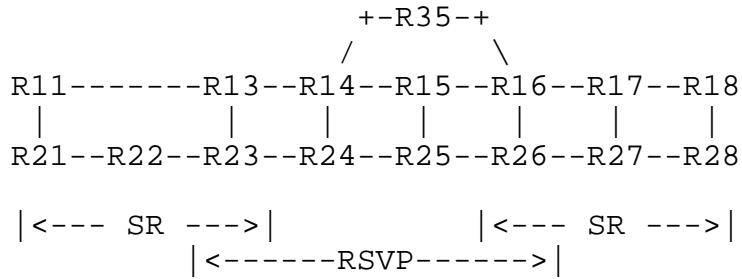
          Figure 2: Example network with segment routing and RSVP

   Figure 2 shows a network that uses both segment routing and RSVP.
   Segment routing is enabled on R11, R13, R21, R22, R23, R16, R17, R18,
   R26, R27, and R28. RSVP is enabled on R13, R14, R15, R16, R23, R24,
   R25, R26, and R35. Note that both segment routing and RSVP are
   enabled on R13, R23, R16, and R26. R23 uses RSVP to signal an LSP
   from R23 to R16, following the path R23->R24->R25->R26->R16. R23
   uses a Binding advertisement to advertise the following values:

   o label value = 2099

   o FEC = 10.0.0.16/32

   o ERO = (10.0.0.24[strict], 10.0.0.25[strict], 10.0.0.26[strict],
      10.0.0.16[strict])

   R11 receives the Binding advertisement and decides (based on some
   policy) to forward traffic from R11 to R18 along a path that consists
   of the following partial paths:

   o the shortest path from R11 to R23

   o the path from R23 to R16 following the explicit path
      R23->R24->R25->R26->R16

   o the shortest path from R16 to R18

   In order to accomplish this, R11 sends packets to R13 with the label
   stack = <1023,2099,1018>. Label 1023 corresponds to the node-SID
   label for R23, and thus results in forwarding along the shortest path
   from R11 to R23. The packets will arrive at R23 with label stack =
   <2099,1018>. The top label value of 2099 at R23 will result in
   forwarding of packets along the path R23->R24->R25->R26->R16 using
   the label SWAP operations signalled by RSVP for this LSP. With
   penultimate hop popping, the packets will arrive at R16 with label
   stack = <1018>. Label 1018 corresponds to the node-SID label for

R18, and thus results in forwarding along the shortest path from R16 to R18.

Note that the scenario described in this example is very similar to the commonly deployed LDP-over-RSVP architecture, with shortest path routing with LDP at the edges and explicit routing with RSVP in the core. However, it is difficult to achieve fine-grained forwarding behavior described in this example using LDP-over-RSVP. In an LDP-over-RSVP network, the only way to influence which which LDP/edge traffic gets tunnelled over which RSVP LSPs is to advertise the RSVP LSPs as forwarding adjacencies (FAs) and tune the IGP metrics of the FAs. It may be difficult or impossible to achieve the desired mapping of LDP/edge traffic over RSVP LSPs using by tuning the metrics of FAs.

Instead, with the SR-and-RSVP architecture described above, it is possible to achieve an arbitrary mapping of edge traffic to core RSVP LSPs using a maximum label stack depth of 3, assuming shortest path forwarding between edge and core nodes via node-SIDs.

One could also build label stacks using adjacency labels advertised by SR-capable routers in the edge networks in order to forward traffic along non-shortest paths in the edge networks. More explicit control of forwarding paths in the edge networks would come at the expense of deeper label stacks.

4.1. Advertising a backup ERO

Continuing with the example of Figure 2, it may also be desirable for R23 to provide information about backup paths that may be used in the event of a failure affecting the primary path. For example, assume that R23 has signalled a primary LSP along the path R23->R24->R25->R26->R16 and a backup LSP along the path R23->R13->R14->R35->R16. R23 uses a Binding advertisement to advertise the following values:

o label value = 2099

o FEC = 10.0.0.16/32

o ERO = (10.0.0.24[strict], 10.0.0.25[strict], 10.0.0.26[strict], 10.0.0.16[strict])

o backup ERO = (10.0.0.13[strict], 10.0.0.14[strict], 10.0.0.35[strict], 10.0.0.16[strict])

R11 may use the information about the backup path in this Binding advertisement to decide whether or not to construct SR label stacks

that use this RSVP LSP. For example, R11 may have requirement avoid
forwarding traffic over primary or backup paths that include R35.

5. Advertising an LDP LSP

```
              R31--R32--R33--R34--R35--R36

              |<------ SR ----->|
                          |<--- LDP --->|
```
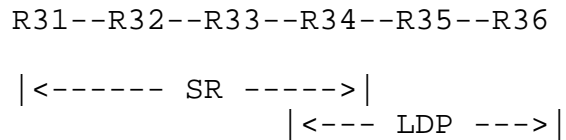
             Figure 3: Example network with segment routing and LDP

Figure 3 shows a network that uses both segment routing and LDP.
Segment routing is enabled on R31-34, while LDP is enabled on R34-36.
Note that both segment routing and LDP are enabled on R34. Also note
that LDP is NOT enabled on R31, R32 and R33. R34 has received a
label mapping for FEC=10.0.0.36/32 from R35 using LDP, corresponding
to an LDP LSP from R34 to R36 along the shortest path. R34 uses a
Binding advertisement to advertise the following values:

o label value = 2088

o FEC = 10.0.0.36/32

After receiving this Binding advertisement, R31 can forward traffic
to R36 by sending packets to next-hop R32 with label stack =
<1034,2088>. The packets will arrive at R34 with label stack =
<2088>. The top label value of 2088 at R34 will result in forwarding
of packets along the shortest path to R36 using the label SWAP
operations for FEC = 10.0.0.36/32 signalled by LDP.

Now we look at what is needed to forward labeled traffic from R36 in
the LDP-only domain to R31 in the SR-only domain. R34 can get a
packet to R31 using R31's node-SID label (1031). In order for R34 to
apply label=1031 to packets in FEC 10.0.0.31, packets need to arrive
at R34 with a label corresponding to FEC 10.0.0.31. Therefore R34
should not use penultimate hop popping when it distributes a label to
the LDP-only domain.

The routers in the LDP-only domain (R34, R35, and R36) advertise
label mappings for FEC 10.0.0.31/32 using LDP. This corresponds to
normal LDP behavior based on [RFC5036] since R34, R35, and R36 each
has an entry in its routing table for 10.0.0.31/32, and in the
absence of segment routing, R34 is an egress LSR with respect to FEC
10.0.0.31/32. This will allow packets to travel from R36 to R34
using LDP labels. Via LDP, R34 advertises the label value 3154 for
FEC 10.0.0.31/32 (instead of implicit or explicit null). Packets

from R36 arrive at R34 with label=3154. R34 recognizes that
label=3154 corresponds to FEC 10.0.0.31/32, so it swaps the label
with outgoing label=1031 (the node-SID for R31), and forwards the
packet to next-hop R33. The packet will ultimately be delivered to
R31 over the shortest path in the SR-only network using the node-SID.

Note that [I-D.filsfils-spring-segment-routing-ldp-interop] describes
a different method (utilizing a Segment Routing Mapping Server) to
allow SR-enabled nodes to send traffic to LDP nodes that do not
support SR.

6. Advertising a BGP labeled-unicast LSP

```
                 region 1     |     region 2     |    region 3

        R71-------R72-----R81-------R82-----R91-------R92
         |         |       |         |       |         |
        R73--R74--R75-----R83--R84--R85-----R93--R94--R95

        |<---- SR ---->|<----- LDP ----->|<---- LDP --->|


               ^       ^ ^       ^ ^       ^ ^         ^
               \___/   \_____/   \___/   \_____/
              BGP-LU   BGP-LU    BGP-LU   BGP-LU
```
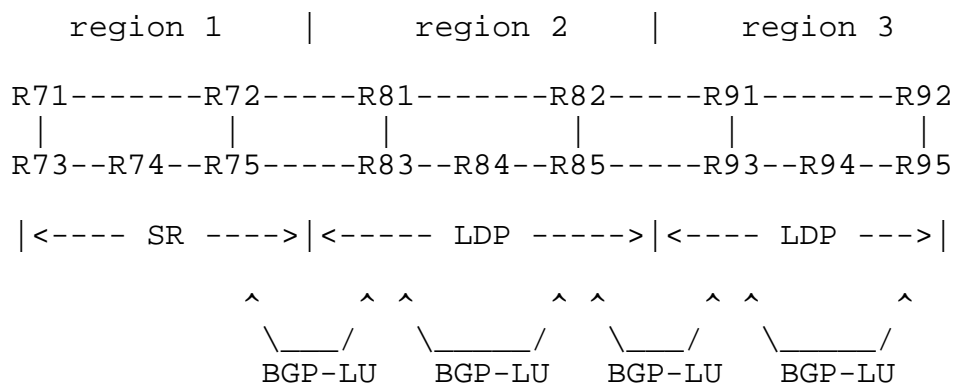
          Figure 4: Example network with segment routing and BGP labeled
                               unicast

Figure 4 shows a network that uses segment routing together with BGP
labeled-unicast (BGP-LU) [RFC3107] . In this example, segment
routing is enabled on R71-75 (region 1). LDP is enabled on R81-84
(region 2) and R91-95 (region 3). In addition, BGP-LU sessions exist
between R75 and R83, R83 and R85, R85 and R93, and R93 and R95. Via
LDP, R93 learns a label value of 3009 from R94 corresponding to FEC
10.0.0.95/32. Via its BGP-LU session with R85, R93 advertises a
label value of 4021 corresponding to prefix 10.0.0.95/32, with a
next-hop of 10.0.0.93. Via its BGP-LU session with R83, R85
advertises a label value of 4031 corresponding to prefix
10.0.0.95/32, with a next-hop of 10.0.0.85. Via its BGP-LU session
with R75, R83 advertises a label value of 4041 corresponding to
prefix 10.0.0.95/32, with a next-hop of 10.0.0.83. R75 uses a
segment routing Binding advertisement to advertise the following
values:

o label value = 2077

o FEC = 10.0.0.95/32

o   ERO = (10.0.0.83[strict])

In this example, R75 has included an ERO list with a single element corresponding to the directly connected next-hop from R75 to R84, which is all that R75 knows about the path based on the BGP-LU advertisement. Its inclusion by R75 is optional, but the information may be useful to routers that receive the Binding advertisement.

R71 receives the Binding advertisement via the IGP. In order to send a labeled packet to R95, R71 needs to construct a label stack that causes the packet to arrive at R75 with top label=2077. If R71 wants the packet to take the shortest path from R71 to R75, then it sends the packet to R72 with label stack = <1075,2077>. (Label 1075 is the node-SID for R75 based on the conventions in Section 3.) The packet arrives at R75 with label stack = <2077>. R75 swaps label 2077 with label 4041 and forwards the packet to next-hop R83. R83 swaps label 4031 with label 4021, pushes the LDP label distributed by R84 for FEC=10.0.0.85, and forwards the packet to next-hop R84. The packet arrives at R85 with label 4031 exposes, so R85 swaps label 4031 with label 4021, and forwards the packet to next-hop R93. Finally R93 swaps label 4021 with label 3009 and forwards the packet to next-hop R94, allowing the packet to be delivered to R95 via LDP label operations.

If instead R71 wants the packet to take the path R71->R73->R74->R75, then it would send the packet to R73 with two adjacency labels corresponding to the links between R73 and R74 and between R74 and R75, followed by the label 2077.

The above description accounts for sending labeled packets from a source in a segment routing region to a destination in another region using paths established by BGP-LU. Since the example is not symmetric with respect to source and destination, for completeness, we illustrate how traffic to send traffic from another region to a destination in a segment routing region using LSPs established by BGP-LU, which in this example corresponds to sending a packet from R95 to R71.

R75 knows that it can send a packet to R71 by imposing a node-SID label of 1071. Via its BGP-LU session with R83, R75 advertises a label value of 4052 for prefix 10.0.0.71/32, with a next-hop of 10.0.0.75. The string of BGP-LU sessions from R83 to R85 to R93 to R95 advertise label bindings for prefix 10.0.0.71/32 such that R95 can send a packet to R93 with the appropriate BGP-LU label that the packet will arrive at R75 with label 4052 exposed as the top label. When R75 receives this packet, it swaps label 4052 with label 1071 and forwards the packet to next-hop R72, resulting in the packet being delivered to R71.

7. Advertising a static LSP

```
                              +-R54-+
                             /       \
                        +-R51  AS5  R53--+
                       /     \       /     \
                      /       +-R52-+       R71---R74
       R41----R42----R43                    | AS7 |
       |             | \     +-R66-+        R72---R73
       |             |  \   /       \       /
       |     AS4     |   +-R61       R65--+
       |             |     |   AS6   |
       |             |     |         |
       R44----R45----R46----R62        R64
                            \         /
                             +-R63-+


             |<---- SR ----->|
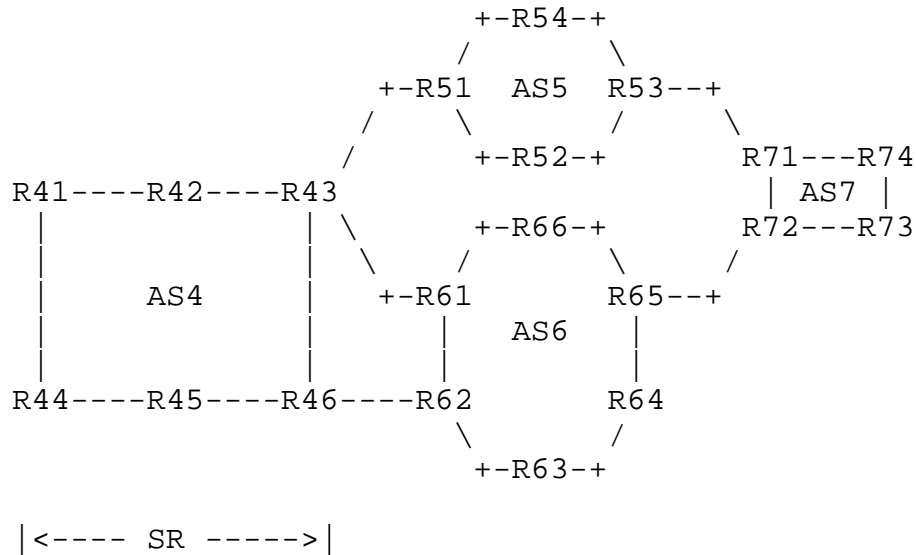```

         Figure 5: Example network using segment routing extensions to
                          advertise a static lsp


   In Figure 5, each grouping of routers (R41-46, R51-45, R61-66, and
   R71-74) represents a different a different autonomous system (AS
   4,5,6, and 7 respectively).  Segment routing is enabled on R41-46.
   R43 has two interfaces that connect to routers outside of its AS.
   One egress interface connects to R51 while another connects to R61.
   It is desirable for routers in AS4 to be able to send traffic to R43
   with a label stack that indicates the interface that R43 should use
   to send the traffic.  This can be accomplished by having R43
   advertise label bindings for one-hop static LSPs corresponding to
   each of its egress interfaces.  In order advertise the egress
   interface connected to R51, R43 uses a segment routing Binding
   advertisement to advertise the following values:

   o  label value = 2033

   o  FEC = 10.0.0.51/32

   o  ERO = (10.0.0.51[strict])

   In order advertise the egress interface connected to R61, R43 uses a
   segment routing Binding advertisement to advertise the following
   values:

o  label value = 2034

o  FEC = 10.0.0.61/32

o  ERO = (10.0.0.61[strict])

R41 receives the Binding advertisements via the IGP.  In order to
send a packet out R43's interface to R51, R41 constructs a packet
with label stack = <1043,2033> and sends it to R42.  (Label 1043 is
the node-SID for R43 based on the conventions in Section 3.)  The
packet arrives at R43 with label stack = <2033>.  R43 will POP label
2033 and send the unlabeled packet out the interface to R51.
Similarly in order to send a packet out R43's interface to R56, R41
constructs a packet with label stack = <1043,2034> and sends it to
R42.

8.  Advertising a context label for egress node protection


```
                          /- RR -\
                         /   \
                PE3---P7----------P8---PE5
                / | | \
            CE1 | | CE2
                \ | | /
                PE4---P9---P10---P11---PE6

            |<--------- LDP ---------->|
```
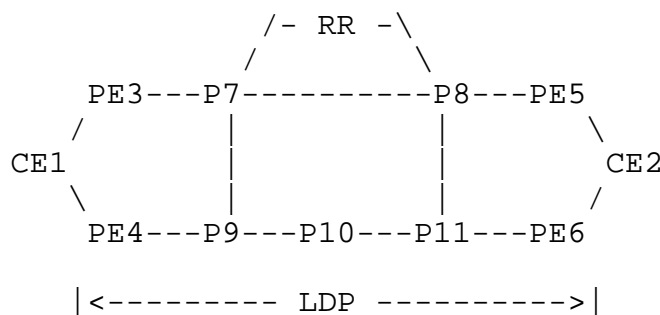

        Figure 6: Example network using segment routing extensions to
             advertise a context label for egress node protection

   [I-D.minto-2547-egress-node-fast-protection] describes a mechanism to
   provide fast protection of RFC 2547/4364 based VPN traffic against
   egress PE failure.  In the example in Figure 6, CE1 and CE2 are each
   dual-homed to two different PEs and belong to the VPN-A.  In the
   absence of a failure, traffic travels from CE1 to CE2 over the path
   CE1->PE3->P7->P8->PE5->CE2.  Using the mechanism described in
   [I-D.minto-2547-egress-node-fast-protection], upon the failure of
   PE5, the point of local repair (P8) can use a loop-free alternate
   (LFA) to divert the traffic destined for protected PE (PE5) to a
   Protector function co-located with an alternate egress PE for CE2
   (PE6).  Before forwarding the traffic to PE6, P8 pushes a context
   label associated with PE5 (the protected PE).  This context label
   allows PE6 to interpret the VPN label in the context of PE5 VPN label
   advertisements, since the VPN label on these packets was originally

imposed by the ingress PE (PE3) based on the assumption that they would be delivered to PE5.

We assume in this example that there is a single context-identifier corresponding to the protected PE (PE5) with a value of 10.1.1.5. The prefix 10.1.1.5/32 is advertised in the IGP and in LDP by PE5. PE5 advertises VPN-IP prefixes via BGP with next-hop = 10.1.1.5. This causes VPN traffic from PE3 to PE5 will take an LDP transport tunnel corresponding to FEC 10.1.1.5/32.

PE6 advertises a context label for context-identifier 10.1.1.5 using a segment routing Binding advertisement with the following values:

o  label value = 2066

o  FEC = 10.1.1.5/32

o  Mirror context = TRUE

When P8 receives this Binding advertisement via the IGP, it creates a forwarding table entry for LDP traffic in FEC 10.1.1.5 that will be activated immediately when the link to PE5 fails.  This behavior is triggered by setting Mirror context = TRUE in the advertisement. This backup forwarding table entry uses a loop-free alternate (LFA) or remote LFA to send traffic to PE6 (using FEC 10.0.0.6 which is the router-id of PE6) along a path that avoids passing through PE5. Importantly, the backup forwarding table entry pushes label 2066 into the packet before applying any labels associated with the repair path to PE6.

When the link to PE5 fails and P8 activates the backup forwarding table entry for LDP traffic in FEC 10.1.1.5, that traffic will be diverted to PE6.  The packets will arrive at PE6 with top label = 2066, followed by the VPN label advertised by PE5.  PE6 pops label 2066, and interprets the next label as a VPN label advertised by PE5. PE6 has been listening in on PE5s BGP advertisements, so it knows the mapping between a given VPN label advertised by PE5 and the actual VPN.

9.  IANA Considerations

This document introduces no new IANA Considerations.

10.  Management Considerations

TBD

11.  Security Considerations

   TBD

12.  Acknowledgements

13.  References

13.1.  Normative References

   [RFC3107]  Rekhter, Y. and E. Rosen, "Carrying Label Information in
              BGP-4", RFC 3107, May 2001.

   [RFC5036]  Andersson, L., Minei, I., and B. Thomas, "LDP
              Specification", RFC 5036, October 2007.

13.2.  Informative References

   [I-D.filsfils-spring-segment-routing]
              Filsfils, C., Previdi, S., Bashandy, A., Decraene, B.,
              Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R.,
              Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe,
              "Segment Routing Architecture", draft-filsfils-spring-
              segment-routing-04 (work in progress), July 2014.

   [I-D.filsfils-spring-segment-routing-ldp-interop]
              Filsfils, C., Previdi, S., Bashandy, A., Decraene, B.,
              Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R.,
              Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe,
              "Segment Routing interoperability with LDP", draft-
              filsfils-spring-segment-routing-ldp-interop-01 (work in
              progress), April 2014.

   [I-D.ietf-isis-segment-routing-extensions]
              Previdi, S., Filsfils, C., Bashandy, A., Gredler, H.,
              Litkowski, S., Decraene, B., and J. Tantsura, "IS-IS
              Extensions for Segment Routing", draft-ietf-isis-segment-
              routing-extensions-02 (work in progress), June 2014.

   [I-D.ietf-ospf-segment-routing-extensions]
              Psenak, P., Previdi, S., Filsfils, C., Gredler, H.,
              Shakir, R., Henderickx, W., and J. Tantsura, "OSPF
              Extensions for Segment Routing", draft-ietf-ospf-segment-
              routing-extensions-02 (work in progress), August 2014.

   [I-D.minto-2547-egress-node-fast-protection]
             Jeganathan, J., Gredler, H., and B. Decraene, "2547 egress
             PE Fast Failure Protection", draft-minto-2547-egress-node-
             fast-protection-03 (work in progress), July 2014.

   [I-D.sivabalan-pce-segment-routing]
             Sivabalan, S., Medved, J., Filsfils, C., Crabbe, E.,
             Raszuk, R., Lopez, V., and J. Tantsura, "PCEP Extensions
             for Segment Routing", draft-sivabalan-pce-segment-
             routing-03 (work in progress), July 2014.

Authors' Addresses

   Chris Bowers
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA  94089
   US


   Email: cbowers@juniper.net


   Hannes Gredler
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA  94089
   US


   Email: hannes@juniper.net


   Uma Chunduri
   Ericsson Inc.
   300 Holger Way
   San Jose, CA  95134
   US


   Email: uma.chunduri@ericsson.com