                                                             B. Decraene
                                                            S. Litkowski
                                                                  Orange
                                                             M. Horneffer
                                                         Deutsche Telekom
                                                               R. Shakir
                                                      Jive Communications
                                                              J. Tantsura
                                                                 Ericsson
                                                               E. Crabbe
                                                               Individual
                                                         February 1, 2016

Segment Routing with MPLS data plane
draft-ietf-spring-segment-routing-mpls-03

Abstract

   Segment Routing (SR) leverages the source routing paradigm. A node
   steers a packet through a controlled set of instructions, called
   segments, by prepending the packet with an SR header. A segment can
   represent any instruction, topological or service-based. SR allows
   to enforce a flow through any topological path and service chain
   while maintaining per-flow state only at the ingress node to the SR
   domain.

   Segment Routing can be directly applied to the MPLS architecture with
   no change in the forwarding plane. This drafts describes how Segment
   Routing operates on top of the MPLS data plane.

Requirements Language

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in RFC 2119 [RFC2119].

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

Table of Contents

1. Introduction

   The Segment Routing architecture [I-D.ietf-spring-segment-routing]
   can be directly applied to the MPLS architecture with no change in
   the MPLS forwarding plane. This drafts describes how Segment Routing
   operates on top of the MPLS data plane.

   The Segment Routing problem statement is described in
   [I-D.ietf-spring-problem-statement].

   Link State protocol extensions for Segment Routing are described in
   [I-D.ietf-isis-segment-routing-extensions],
   [I-D.ietf-ospf-segment-routing-extensions] and
   [I-D.ietf-ospf-ospfv3-segment-routing-extensions].

2. Illustration

   Segment Routing, applied to the MPLS data plane, offers the ability
   to tunnel services (VPN, VPLS, VPWS) from an ingress PE to an egress
   PE, without any other protocol than ISIS or OSPF
   ([I-D.ietf-isis-segment-routing-extensions] and
   [I-D.ietf-ospf-segment-routing-extensions]). LDP and RSVP-TE
   signaling protocols are not required.

   Note that [I-D.filsfils-spring-segment-routing-ldp-interop] documents
   SR co-existence and interworking with other MPLS signaling protocols,
   if present in the network during a migration, or in case of non-
   homogeneous deployments.

   The operator only needs to allocate one node segment per PE and the
   SR IGP control-plane automatically builds the required MPLS
   forwarding constructs from any PE to any PE.

```
                     P1---P2
                    /        \
          A---CE1---PE1         PE2---CE2---Z
                    \        /
                     P3---P4
```
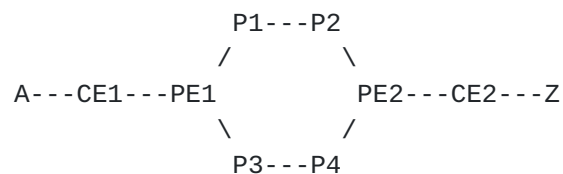
                 Figure 1: IGP-based MPLS Tunneling

   In Figure 1 above, the four nodes A, CE1, CE2 and Z are part of the
   same VPN.

   PE2 advertises (in the IGP) a host address 192.0.2.2/32 with its
   attached node segment 102.

CE2 advertises to PE2 a route to Z. PE2 binds a local label LZ to
that route and propagates the route and its label via MPBGP to PE1
with nhop 192.0.2.2 (PE2 loopback address).

PE1 installs the VPN prefix Z in the appropriate VRF and resolves the
next-hop onto the node segment 102. Upon receiving a packet from A
destined to Z, PE1 pushes two labels onto the packet: the top label
is 102, the bottom label is LZ. 102 identifies the node segment to
PE2 and hence transports the packet along the ECMP-aware shortest-
path to PE2. PE2 then processes the VPN label LZ and forwards the
packet to CE2.

Supporting MPLS services (VPN, VPLS, VPWS) with SR has the following
benefits:

   Simple operation: one single intra-domain protocol to operate: the
   IGP. No need to support IGP synchronization extensions as
   described in [RFC5443] and [RFC6138].

   Excellent scaling: one Node-SID per PE.

3. MPLS Instantiation of Segment Routing

MPLS instantiation of Segment Routing fits in the MPLS architecture
as defined in [RFC3031] both from a control plane and forwarding
plane perspective:

o From a control plane perspective [RFC3031] does not mandate a
   single signaling protocol. Segment Routing proposes to use the
   Link State IGP as its use of information flooding fits very well
   with label stacking on ingress.

o From a forwarding plane perspective, Segment Routing does not
   require any change to the forwarding plane.

When applied to MPLS, a Segment is a LSP and the 20 right-most bits
of the segment are encoded as a label. This implies that, in the
MPLS instantiation, the SID values are allocated within a reduced
20-bit space out of the 32-bit SID space.

The notion of indexed global segment fits the MPLS architecture
[RFC3031] as the absolute value allocated to any segment (global or
local) can be managed by a local allocation process (similarly to
other MPLS signaling protocols).

If present, SR can coexist and interworks with LDP and RSVP
[I-D.filsfils-spring-segment-routing-ldp-interop].

The source routing model described in
[I-D.ietf-spring-segment-routing] is inherited from the ones proposed
by [RFC1940] and [RFC2460]. The source routing model offers the
support for explicit routing capability.

Contrary to RSVP-based explicit routes where tunnel midpoints
maintain states, SR-based explicit routes only require per-flow
states at the ingress edge router where the traffic engineer policy
is applied.

Contrary to RSVP-based explicit routes which consist in non-ECMP
circuits (similar to ATM/FR), SR-based explicit routes can be built
as list of ECMP-aware node segments and hence ECMP-aware traffic
engineering is natively supported by SR.

When Segment Routing is instantiated over the MPLS data plane the
following applies:

o A list of segments is represented as a stack of labels.

o The active segment is the top label.

o The CONTINUE operation is implemented as an MPLS swap operation.
   The outgoing label value is computed as follows:

   * When the same Segment Routing Global Block (SRGB, defined in
      [I-D.ietf-spring-segment-routing] is used throughout the SR
      domain, the outgoing label value is equal to the incoming label
      value.

   * When different SRGBs are used, the outgoing label value is set
      as: [SRGB(next_hop)+index]. If the index can't be applied to
      the SRGB (i.e.: if the index points outside the SRGB of the
      next-hop or the next-hop has not advertised a valid SRGB), then
      no outgoing label value can be computed and the next-hop MUST
      be considered as not supporting the MPLS operations for that
      particular SID.

o The NEXT operation is implemented as an MPLS pop operation. The
   NEXT operation does not require any mapping to an outgoing label
   hence the SRGB is irrelevant for this operation.

o The PUSH operation is implemented as an MPLS push of a label
   stack.

o The Segment Routing Global Block (SRGB) values MUST be greater
   than 15 in order to preserve values 0-15 as defined in [RFC3032].

o As described in [I-D.ietf-spring-segment-routing], using the same
   SRGB on all nodes within the SR domain eases operations and
   troubleshooting and is expected to be a deployment guideline.

In conclusion, there are no changes in the operations of the data-
plane currently used in MPLS networks.

4. IGP Segments Examples

Assuming the network diagram of Figure 2 and the IP address and IGP
Segment allocation of Figure 3, the following examples can be
constructed.

```
                    +--------+
                   /          \
           R1-----R2----------R3-----R8
                  | \        / |
                  |  +--R4--+  |
                  |           |
                  +-----R5-----+
```

                Figure 2: IGP Segments - Illustration

```
+-----------------------------------------------------------+
| IP address allocated by the operator:                     |
|                   192.0.2.1/32 as a loopback of R1         |
|                   192.0.2.2/32 as a loopback of R2         |
|                   192.0.2.3/32 as a loopback of R3         |
|                   192.0.2.4/32 as a loopback of R4         |
|                   192.0.2.5/32 as a loopback of R5         |
|                   192.0.2.8/32 as a loopback of R8         |
|             198.51.100.9/32 as an anycast loopback of R4   |
|             198.51.100.9/32 as an anycast loopback of R5   |
|                                                           |
| SRGB defined by the operator as 1000-5000                 |
|                                                           |
| Global IGP SID allocated by the operator:                 |
|                   1001 allocated to 192.0.2.1/32          |
|                   1002 allocated to 192.0.2.2/32          |
|                   1003 allocated to 192.0.2.3/32          |
|                   1004 allocated to 192.0.2.4/32          |
|                   1008 allocated to 192.0.2.8/32          |
|                   2009 allocated to 198.51.100.9/32       |
|                                                           |
| Local IGP SID allocated dynamically by R2                 |
|                   for its "north" adjacency to R3: 9001   |
|                   for its "north" adjacency to R3: 9003   |
|                   for its "south" adjacency to R3: 9002   |
|                   for its "south" adjacency to R3: 9003   |
+-----------------------------------------------------------+
```
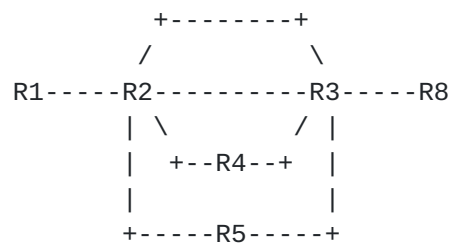
       Figure 3: IGP Address and Segment Allocation - Illustration

4.1. Example 1

   R1 may send a packet P1 to R8 simply by pushing an SR header with
   segment list {1008}.

   1008 is a global IGP segment attached to the IP prefix 192.0.2.8/32.
   Its semantic is global within the IGP domain: any router forwards a
   packet received with active segment 1008 to the next-hop along the
   ECMP-aware shortest-path to the related prefix.

   In conclusion, the path followed by P1 is R1-R2--R3-R8. The ECMP-
   awareness ensures that the traffic be load-shared between any ECMP
   path, in this case the two north and south links between R2 and R3.

4.2. Example 2

   R1 may send a packet P2 to R8 by pushing an SR header with segment
   list {1002, 9001, 1008}.

   1002 is a global IGP segment attached to the IP prefix 192.0.2.2/32.
   Its semantic is global within the IGP domain: any router forwards a
   packet received with active segment 1002 to the next-hop along the
   shortest-path to the related prefix.

   9001 is a local IGP segment attached by node R2 to its north link to
   R3. Its semantic is local to node R2: R2 switches a packet received
   with active segment 9001 towards the north link to R3.

   In conclusion, the path followed by P2 is R1-R2-north-link-R3-R8.

4.3. Example 3

   R1 may send a packet P3 along the same exact path as P1 using a
   different segment list {1002, 9003, 1008}.

   9003 is a local IGP segment attached by node R2 to both its north and
   south links to R3. Its semantic is local to node R2: R2 switches a
   packet received with active segment 9003 towards either the north or
   south links to R3 (e.g. per-flow loadbalancing decision).

   In conclusion, the path followed by P3 is R1-R2-any-link-R3-R8.

4.4. Example 4

   R1 may send a packet P4 to R8 while avoiding the links between R2 and
   R3 by pushing an SR header with segment list {1004, 1008}.

   1004 is a global IGP segment attached to the IP prefix 192.0.2.4/32.
   Its semantic is global within the IGP domain: any router forwards a
   packet received with active segment 1004 to the next-hop along the
   shortest-path to the related prefix.

   In conclusion, the path followed by P4 is R1-R2-R4-R3-R8.

4.5. Example 5

   R1 may send a packet P5 to R8 while avoiding the links between R2 and
   R3 while still benefitting from all the remaining shortest paths (via
   R4 and R5) by pushing an SR header with segment list {2009, 1008}.

   2009 is a global IGP segment attached to the anycast IP prefix
   198.51.100.9/32. Its semantic is global within the IGP domain: any

router forwards a packet received with active segment 2009 to the
next-hop along the shortest-path to the related prefix.

In conclusion, the path followed by P5 is either R1-R2-R4-R3-R8 or
R1-R2-R5-R3-R8 .

5. Other Examples of MPLS Segments

In addition to the IGP segments previously described, the SPRING
source routing policy applied to MPLS can include MPLS LSP's signaled
by LDP, RSVPTE and BGP. The list of examples is non exhaustive.
Other form of segments combination can be instantiated through
Segment Routing (e.g.: RSVP LSPs combined with LDP or IGP or BGP
LSPs).

5.1. LDP LSP segment combined with IGP segments

The example illustrates a segment-routing policy including IGP
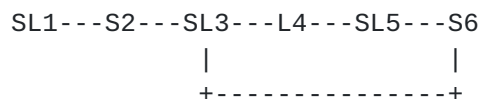segments and LDP LSP segments.

```
            SL1---S2---SL3---L4---SL5---S6
                        |             |
                        +-------------+
```

          Figure 4: LDP LSP segment combined with IGP segments

We assume that:

o All links have an IGP cost of 1 except SL3-S6 link which has cost
   2.

o All nodes are in the same IGP area.

o Nodes SL1, S2, SL3, SL5 and S6 are IGP-SR capable.

o SL3 and S6 have, respectively, index 3 and 6 assigned to them.

o All SR nodes have the same SRGB consisting of: [1000, 1999]

o SL1, SL3, L4 and SL5 are LDP capable.

o SL1 has a directed LDP session with SL3 and is able to retrieve
   the SL3 local LDP mapping for FEC SL5: 35

o The following source-routed policy is defined in SL1 for the
   traffic destined to S6: use path SL1-S2-SL3-L4-SL5-S6 (instead of
   shortest-path SL1-S2-SL3-S6).

This is realized by programming the following segment-routing policy
at S1: for traffic destined to S6, push the ordered segment list:
{1003, 35, 1006}, where:

o 1003 gets the packets from S1 to SL3 via S2.

o 35 gets the packets from SL3 to SL5 via L4.

o 1006 gets the packets from SL5 to S6.

The above allows to steer the traffic into path SL1-S2-SL3-L4-SL5-S6
instead of the shortest path SL1-S2-SL3-S6.

5.2. RSVP-TE LSP segment combined with IGP segments

The example illustrates a segment-routing policy including IGP
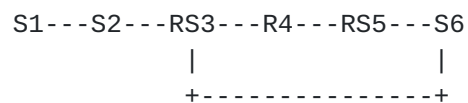segments and RSVP-TE LSP segments.

```
          S1---S2---RS3---R4---RS5---S6
                     |              |
                     +--------------+
```

        Figure 5: RSVP-TE LSP segment combined with IGP segments

We assume that:

o All links have an IGP cost of 1 except link RS3-S6 which has cost
   2.

o All nodes are IGP-SR capable except R4.

o RS3 and R6 have, respectively, index 3 and 6 assigned to them.

o All SR nodes have the same SRGB consisting of: [1000, 1999]

o RS3, R4 and RS5 are RSVP-TE capable.

o An RSVP-TE LSP has been provisioned from RS3 to RS5 via R4.

o RS3 allocates a binding SID (with value of 135) for this RSVP-TE
   LSP and signals it in the igp.

o The following source-routed policy is defined at S1 for the
   traffic destined to S6: use path S1-S2-RS3-R4-RS5-S6 instead of
   shortest-path S1-S2-RS3-S6.

This is realized by programming the following segment-routing policy
at S1: - for traffic destined to S6, push the ordered segment list:
{1003, 135, 1006}, where:

o 1003 gets the packets from S1 to RS3 via S2.

o 135 gets the packets from RS3 into the RSVP-TE LSP to RS5 via R4.

o 1006 gets the packets from RS5 to S6.

The above allows to steer the traffic into path S1-S2-RS3-R4-RS5-S6
instead of the shortest path S1-S2-RS3-S6.

[6](#). Segment List History

In the abstract SR routing model [[I-D.ietf-spring-segment-routing](#)],
any node N along the journey of the packet is able to determine where
the packet P entered the SR domain and where it will exit. The
intermediate node is also able to determine the paths from the
ingress edge router to itself, and from itself to the egress edge
router.

In the MPLS instantiation, as the packet travels through the SR
domain, the stack is depleted and the segment list history is
gradually lost.

[7](#). IANA Considerations

This document doesn't introduce any codepoint.

[8](#). Manageability Considerations

TBD

[9](#). Security Considerations

TBD

[10](#). Contributors

The following contributors have substantially helped the definition
and editing of the content of this document:

Wim Henderickx
Email: wim.henderickx@alcatel-lucent.com

Igor Milojevic
Email: milojevicigor@gmail.com

Saku Ytti
Email: saku@ytti.fi

[11](). Acknowledgements

[12](). References

[12.1](). Normative References

[RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
            Requirement Levels", [BCP 14](), [RFC 2119](),
            DOI 10.17487/RFC2119, March 1997,
            <[http://www.rfc-editor.org/info/rfc2119]()>.

[RFC2460]   Deering, S. and R. Hinden, "Internet Protocol, Version 6
            (IPv6) Specification", [RFC 2460](), DOI 10.17487/RFC2460,
            December 1998, <[http://www.rfc-editor.org/info/rfc2460]()>.

[RFC3031]   Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
            Label Switching Architecture", [RFC 3031](),
            DOI 10.17487/RFC3031, January 2001,
            <[http://www.rfc-editor.org/info/rfc3031]()>.

[RFC3032]   Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y.,
            Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack
            Encoding", [RFC 3032](), DOI 10.17487/RFC3032, January 2001,
            <[http://www.rfc-editor.org/info/rfc3032]()>.

[12.2](). Informative References

[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-03]() (work in
            progress), March 2015.

[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-06]() (work in progress), December 2015.

   [I-D.ietf-ospf-ospfv3-segment-routing-extensions]
              Psenak, P., Previdi, S., Filsfils, C., Gredler, H.,
              Shakir, R., Henderickx, W., and J. Tantsura, "OSPFv3
              Extensions for Segment Routing", draft-ietf-ospf-ospfv3-
              segment-routing-extensions-04 (work in progress), December
              2015.

   [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-06 (work in progress), December 2015.

   [I-D.ietf-spring-problem-statement]
              Previdi, S., Filsfils, C., Decraene, B., Litkowski, S.,
              Horneffer, M., and r. rjs@rob.sh, "SPRING Problem
              Statement and Requirements", draft-ietf-spring-problem-
              statement-06 (work in progress), December 2015.

   [I-D.ietf-spring-segment-routing]
              Filsfils, C., Previdi, S., Decraene, B., Litkowski, S.,
              and r. rjs@rob.sh, "Segment Routing Architecture", draft-
              ietf-spring-segment-routing-07 (work in progress),
              December 2015.

   [RFC1940] Estrin, D., Li, T., Rekhter, Y., Varadhan, K., and D.
              Zappala, "Source Demand Routing: Packet Format and
              Forwarding Specification (Version 1)", RFC 1940,
              DOI 10.17487/RFC1940, May 1996,
              <http://www.rfc-editor.org/info/rfc1940>.

   [RFC5443] Jork, M., Atlas, A., and L. Fang, "LDP IGP
              Synchronization", RFC 5443, DOI 10.17487/RFC5443, March
              2009, <http://www.rfc-editor.org/info/rfc5443>.

   [RFC6138] Kini, S., Ed. and W. Lu, Ed., "LDP IGP Synchronization for
              Broadcast Networks", RFC 6138, DOI 10.17487/RFC6138,
              February 2011, <http://www.rfc-editor.org/info/rfc6138>.

Authors' Addresses

   Clarence Filsfils (editor)
   Cisco Systems, Inc.
   Brussels
   BE

   Email: cfilsfil@cisco.com

Stefano Previdi (editor)
Cisco Systems, Inc.
Via Del Serafico, 200
Rome   00142
Italy

Email: sprevidi@cisco.com

Ahmed Bashandy
Cisco Systems, Inc.
170, West Tasman Drive
San Jose, CA   95134
US

Email: bashandy@cisco.com

Bruno Decraene
Orange
FR

Email: bruno.decraene@orange.com

Stephane Litkowski
Orange
FR

Email: stephane.litkowski@orange.com

Martin Horneffer
Deutsche Telekom
Hammer Str. 216-226
Muenster   48153
DE

Email: Martin.Horneffer@telekom.de

Rob Shakir
Jive Communications, Inc.
1275 West 1600 North, Suite 100
Orem, UT   84057

Email: rjs@rob.sh

Jeff Tantsura
Ericsson
300 Holger Way
San Jose, CA 95134
US

Email: Jeff.Tantsura@ericsson.com

Edward Crabbe
Individual

Email: edward.crabbe@gmail.com