Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 22, 2017

C. Filsfils, Ed.
S. Previdi, Ed.
A. Bashandy
Cisco Systems, Inc.
B. Decraene
S. Litkowski
Orange
R. Shakir
Google

June 20, 2017

Segment Routing with MPLS data plane
draft-ietf-spring-segment-routing-mpls-09

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. In the MPLS dataplane, the SR header is instantiated through a label stack. A segment can represent any instruction, topological or service-based. Additional segments can be defined in the future. SR allows to enforce a flow through any topological path and/or 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

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 [RFC7855].

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].

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.ietf-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.

2. 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 makes use of Link State IGPs since their flooding mechanism 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 SID 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, defined in [I-D.ietf-spring-segment-routing], 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).

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 (defined in [I-D.ietf-spring-segment-routing]) 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.

  * The index and the SRGB may be learned through different means. Obviously, the SRGB MUST be the one the index is related to.

- o The NEXT operation (defined in [I-D.ietf-spring-segment-routing]) 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 (defined in [I-D.ietf-spring-segment-routing]) 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.

Note that the kind of deployment of Segment Routing may affect the depth of the MPLS label stack. As every segment in the list is represented by an additional MPLS label, the length of the segment list directly correlates to the depth of the label stack. Implementing a long path with many explicit hops as a segment list may thus yield a deep label stack that would need to be pushed at the head of the SR tunnel.

However, many use cases would need very few segments in the list. This is especially true when taking good advantage of the ECMP aware routing within each segment. In fact, most use cases need just one additional segment and thus lead to a similar label stack depth as e.g. RSVP-based routing.

Moreover, the use of the binding segment as specified in [I-D.ietf-spring-segment-routing], also allows to substantially reduce the length of the segment list and hence the depth of the label stack.

Nodes will often have limits with respect to the label depth supported for a PUSH operation. Two ways can be seen to deal with this limitation:

When Segment Routing tunnels are computed by a centralized controller, the controller can consider the Maximum SID depth capability of a node as it may be signaled through routing protocols extensions.

When Segment Routing tunnels are not computed by a centralized controller but derived from an operator designed policy, the operator needs to be aware of the limits of the used nodes and take this into account in the design.

3. IGP Segments Examples

Assuming the network diagram of Figure 1 and the IP address and IGP Segment allocation of Figure 2, the following examples can be constructed.

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

Figure 1: 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 2: IGP Address and Segment Allocation - Illustration

3.1. 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.

3.2. 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.

3.3. 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.

3.4. 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.

3.5. Example 5

R1 may send a packet P5 to R8 while avoiding the links between R2 and R3 while still beneficing 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 .

4. IANA Considerations

This document does not make any request to IANA.

5. Manageability Considerations

This document describes the applicability of Segment Routing over the MPLS data plane. Segment Routing does not introduce any change in the MPLS data plane. Manageability considerations described in [I-D.ietf-spring-segment-routing] applies to the MPLS data plane when used with Segment Routing.

6. Security Considerations

This document does not introduce additional security requirements and mechanisms other than the ones described in [I-D.ietf-spring-segment-routing].

7. Contributors

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

Martin Horneffer
Deutsche Telekom
Email: Martin.Horneffer@telekom.de

Wim Henderickx
Nokia
Email: wim.henderickx@nokia.com

Jeff Tantsura
Email: jefftant@gmail.com

Edward Crabbe
Email: edward.crabbe@gmail.com

Igor Milojevic
Email: milojevicigor@gmail.com

Saku Ytti
Email: saku@ytti.fi

8. Acknowledgements

The authors would like to thank Les Ginsberg and Shah Himanshu for their comments on this document.

9. References

9.1. Normative References

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-11 (work in progress), February 2017.

[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>.

[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>.

9.2. Informative References

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., Decraene, B., and j. jefftant@gmail.com, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-13 (work in progress), June 2017.

[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-09 (work in progress), March 2017.

[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-16 (work in progress), May 2017.

[I-D.ietf-spring-segment-routing-ldp-interop]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., and S. Litkowski, "Segment Routing interworking with LDP", draft-ietf-spring-segment-routing-ldp-interop-08 (work in progress), June 2017.

[RFC7855] Previdi, S., Ed., Filsfils, C., Ed., Decraene, B., Litkowski, S., Horneffer, M., and R. Shakir, "Source Packet Routing in Networking (SPRING) Problem Statement and Requirements", RFC 7855, DOI 10.17487/RFC7855, May 2016, <http://www.rfc-editor.org/info/rfc7855>.

Authors' Addresses

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

Email: cfilsfil@cisco.com

Stefano Previdi (editor)
Cisco Systems, Inc.
Italy

Email: stefano@previdi.net

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

Rob Shakir
Google

Email: robjs@google.com