Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 4, 2019

X. Xu
Alibaba, Inc
S. Bryant
Huawei
A. Farrel
Old Dog Consulting
S. Hassan
Cisco
W. Henderickx
Nokia
Z. Li
Huawei

March 3, 2019

SR-MPLS over IP
draft-ietf-mpls-sr-over-ip-03

Abstract

MPLS Segment Routing (SR-MPLS) is an MPLS data plane-based source routing paradigm in which the sender of a packet is allowed to partially or completely specify the route the packet takes through the network by imposing stacked MPLS labels on the packet. SR-MPLS could be leveraged to realize a source routing mechanism across MPLS, IPv4, and IPv6 data planes by using an MPLS label stack as a source routing instruction set while preserving backward compatibility with SR-MPLS.

This document describes how SR-MPLS capable routers and IP-only routers can seamlessly co-exist and interoperate through the use of SR-MPLS label stacks and IP encapsulation/tunneling such as MPLS-in-UDP as defined in RFC 7510.

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 https://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 September 4, 2019.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

MPLS Segment Routing (SR-MPLS) [I-D.ietf-spring-segment-routing-mpls] is an MPLS data plane-based source routing paradigm in which the sender of a packet is allowed to partially or completely specify the route the packet takes through the network by imposing stacked MPLS labels on the packet. SR-MPLS uses an MPLS label stack to encode a source routing instruction set. This can be used to realize a source routing mechanism that can operate across MPLS, IPv4, and IPv6 data planes. This approach preserves backward compatibility with SR-MPLS. More specifically, the source routing instruction set information

contained in a source routed packet could be uniformly encoded as an MPLS label stack no matter whether the underlay is IPv4, IPv6, or MPLS.

This document describes how SR-MPLS capable routers and IP-only routers can seamlessly co-exist and interoperate through the use of SR-MPLS label stacks and IP encapsulation/tunneling such as MPLS-in-UDP [RFC7510].

Section 2 describes various use cases for the tunneling SR-MPLS over IP. Section 3 describes a typical application scenario and how the packet forwarding happens.

1.1. Terminology

This memo makes use of the terms defined in [RFC3031] and [I-D.ietf-spring-segment-routing-mpls].

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Use Cases

Tunneling SR-MPLS using IPv4 and/or IPv6 tunnels is useful at least in the following use cases:

- Incremental deployment of the SR-MPLS technology may be facilitated by tunneling SR-MPLS packets across parts of a network that are not SR-MPLS enabled using an IP tunneling mechanism such as MPLS-in-UDP [RFC7510]. The tunnel selected MUST have its remote end point (destination) address equal to the address of the next SR-MPLS capable node along the path (i.e., the egress of the active node segment). This is shown in Figure 1.

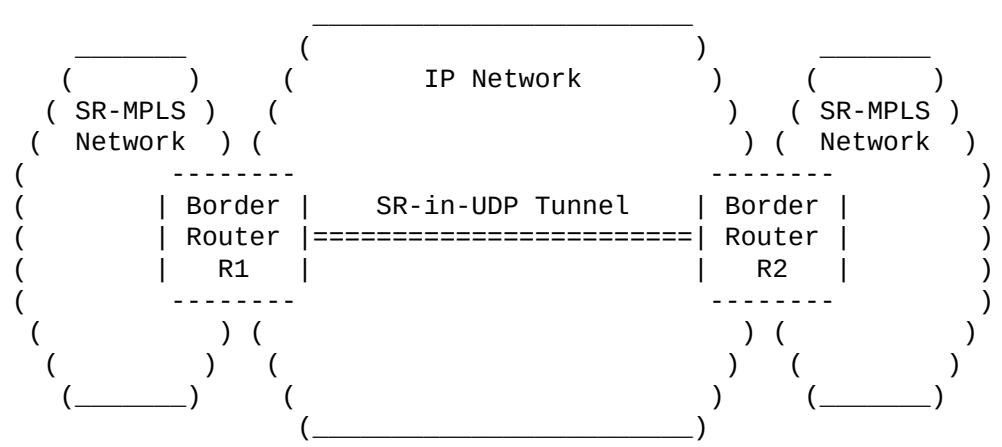

Figure 1: SR-MPLS in UDP to Tunnel Between SR-MPLS Sites

- If encoding of entropy ([RFC6790] is desired, IP tunneling mechanisms that allow encoding of entropy, such as MPLS-in-UDP encapsulation [RFC7510] where the source port of the UDP header is used as an entropy field, may be used to maximize the utilization of ECMP and/or LAG, especially when it is difficult to make use of the entropy label mechanism. Refer to [I-D.ietf-mpls-spring-entropy-label]) for more discussion about using entropy labels in SR-MPLS.

- Tunneling MPLS into IP provides a technology that enables SR in an IPv4 and/or IPv6 network where the routers do not support SRv6 capabilities [I-D.ietf-6man-segment-routing-header] and where MPLS forwarding is not an option. This is shown in Figure 2.

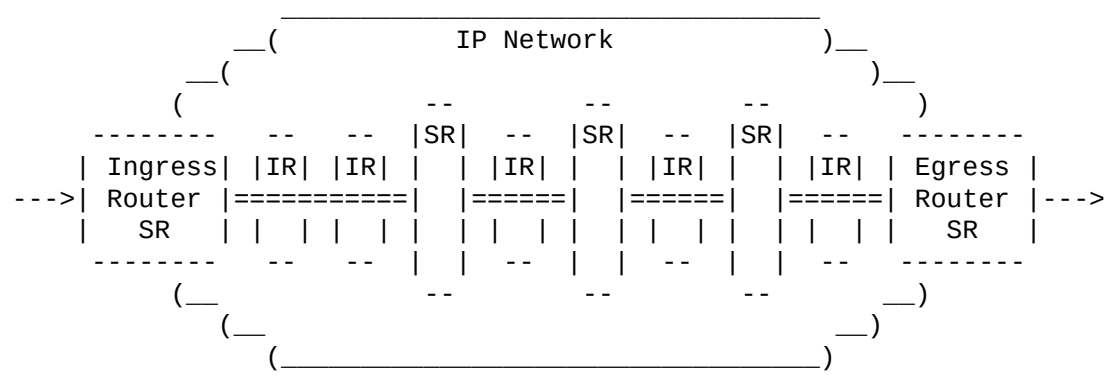

Figure 2: SR-MPLS Enabled Within an IP Network

3. Procedures of SR-MPLS over IP

This section describes the construction of forwarding information base (FIB) entries and the forwarding behavior that allow the deployment of SR-MPLS when some routers in the network are IP only (i.e., do not support SR-MPLS). Note that the examples in Section 3.1 and Section 3.2 assume that OSPF or ISIS is enabled: in fact, other mechanisms of discovery and advertisement could be used including other routing protocols (such as BGP) or a central controller.

3.1. Forwarding Entry Construction

This sub-section describes the how to construct the forwarding information base (FIB) entry on an SR-MPLS-capable router when some or all of the next-hops along the shortest path towards a prefix Segment Identifier (prefix-SID) are IP-only routers.

Consider router A that receives a labeled packet with top label L(E) that corresponds to the prefix-SID SID(E) of prefix P(E) advertised by router E. Suppose the i-th next-hop router (termed NHi) along the shortest path from router A toward SID(E) is not SR-MPLS capable while both routers A and E are SR-MPLS capable. The following processing steps apply:

- The Segment Routing Global Block (SRGB) is defined in [RFC8402]. Router E is SR-MPLS capable, so it advertises an SRGB as described in [I-D.ietf-ospf-segment-routing-extensions] and [I-D.ietf-isis-segment-routing-extensions].

- When Router E advertises the prefix-SID SID(E) of prefix P(E) it MUST also advertise the encapsulation endpoint and the tunnel type of any tunnel used to reach E. It does this using the mechanisms described in [I-D.ietf-isis-encapsulation-cap] or [I-D.ietf-ospf-encapsulation-cap].

- If A and E are in different IGP areas/levels, then:

  * The OSPF Tunnel Encapsulation TLV [I-D.ietf-ospf-encapsulation-cap] or the ISIS Tunnel Encapsulation sub-TLV [I-D.ietf-isis-encapsulation-cap] is flooded domain-wide.

  * The OSPF SID/label range TLV [I-D.ietf-ospf-segment-routing-extensions] or the ISIS SR-Capabilities Sub-TLV [I-D.ietf-isis-segment-routing-extensions] is advertised domain-wide. This way router A knows the characteristics of the router that originated the advertisement of SID(E) (i.e., router E).

  * When router E advertises the prefix P(E):

    + If router E is running ISIS it uses the extended reachability TLV (TLVs 135, 235, 236, 237) and associates the IPv4/IPv6 or IPv4/IPv6 source router ID sub-TLV(s) [RFC7794].

    + If router E is running OSPF it uses the OSPFv2 Extended Prefix Opaque LSA [RFC7684] and sets the flooding scope to AS-wide.

  * If router E is running ISIS and advertises the ISIS capabilities TLV (TLV 242) [RFC7981], it MUST set the "router-ID" field to a valid value or include an IPV6 TE router-ID sub-TLV (TLV 12), or do both. The "S" bit (flooding scope) of the ISIS capabilities TLV (TLV 242) MUST be set to "1" .

- Router A programs the FIB entry for prefix P(E) corresponding to the SID(E) as follows:

  * If the NP flag in OSPF or the P flag in ISIS is clear:

    pop the top label

* If the NP flag in OSPF or the P flag in ISIS is set:

  swap the top label to a value equal to SID(E) plus the lower bound of the SRGB of E

Once constructed, the FIB can be used to tell a router how to process packets. It encapsulates the packets according to the encapsulation advertised in [I-D.ietf-isis-encapsulation-cap] or [I-D.ietf-ospf-encapsulation-cap]. Then it sends the packets towards the next hop NHi.

3.2. Packet Forwarding Procedures

[RFC7510] specifies an IP-based encapsulation for MPLS, i.e., MPLS-in-UDP. This approach is applicable where IP-based encapsulation for MPLS is required and further fine-grained load balancing of MPLS packets over IP networks over Equal-Cost Multipath (ECMP) and/or Link Aggregation Groups (LAGs) is also required. This section provides details about the forwarding procedure when when UDP encapsulation is adopted for SR-MPLS over IP.

Nodes that are SR-MPLS capable can process SR-MPLS packets. Not all of the nodes in an SR-MPLS domain are SR-MPLS capable. Some nodes may be "legacy routers" that cannot handle SR-MPLS packets but can forward IP packets. An SR-MPLS-capable node MAY advertise its capabilities using the IGP as described in Section 3. There are six types of node in an SR-MPLS domain:

- Domain ingress nodes that receive packets and encapsulate them for transmission across the domain. Those packets may be any payload protocol including native IP packets or packets that are already MPLS encapsulated.

- Legacy transit nodes that are IP routers but that are not SR-MPLS capable (i.e., are not able to perform segment routing).

- Transit nodes that are SR-MPLS capable but that are not identified by a SID in the SID stack.

- Transit nodes that are SR-MPLS capable and need to perform SR-MPLS routing because they are identified by a SID in the SID stack.

- The penultimate SR-MPLS capable node on the path that processes the last SID on the stack on behalf of the domain egress node.

- The domain egress node that forwards the payload packet for ultimate delivery.

3.2.1. Packet Forwarding with Penultimate Hop Popping

The description in this section assumes that the label associated with each prefix-SID is advertised by the owner of the prefix-SID is a Penultimate Hop Popping (PHP) label. That is, the NP flag in OSPF or the P flag in ISIS associated with the prefix SID is not set.

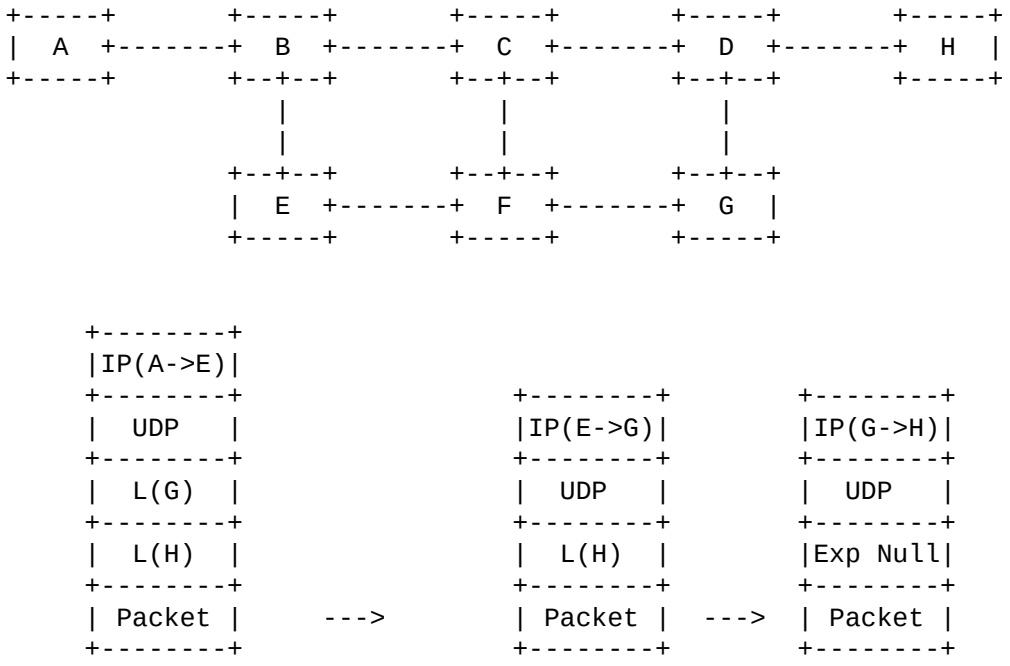

Figure 3: Packet Forwarding Example with PHP

In the example shown in Figure 3, assume that routers A, E, G and H are SR-MPLS-capable while the remaining routers (B, C, D and F) are only capable of forwarding IP packets. Routers A, E, G, and H advertise their Segment Routing related information via IS-IS or OSPF.

Now assume that router A (the Domain ingress) wants to send a packet to router H (the Domain egress) via the explicit path {E->G->H}. Router A will impose an MPLS label stack on the packet that corresponds to that explicit path. Since the next hop toward router E is only IP-capable (B is a legacy transit node), router A replaces the top label (that indicated router E) with a UDP-based tunnel for MPLS (i.e., MPLS-over-UDP [RFC7510]) to router E and then sends the

packet. In other words, router A pops the top label and then encapsulates the MPLS packet in a UDP tunnel to router E.

When the IP-encapsulated MPLS packet arrives at router E (which is an SR-MPLS-capable transit node), router E strips the IP-based tunnel header and then processes the decapsulated MPLS packet. The top label indicates that the packet must be forwarded toward router G. Since the next hop toward router G is only IP-capable, router E replaces the current top label with an MPLS-over-UDP tunnel toward router G and sends it out. That is, router E pops the top label and then encapsulates the MPLS packet in a UDP tunnel to router G.

When the packet arrives at router G, router G will strip the IP-based tunnel header and then process the decapsulated MPLS packet. The top label indicates that the packet must be forwarded toward router H. Since the next hop toward router H is only IP-capable (D is a legacy transit router), router G would replace the current top label with an MPLS-over-UDP tunnel toward router H and send it out. However, since router G reaches the bottom of the label stack (G is the penultimate SR-MPLS capable node on the path) this would leave the original packet that router A wanted to send to router H encapsulated in UDP as if it was MPLS (i.e., with a UDP header and destination port indicating MPLS) even though the original packet could have been any protocol. That is, the final SR-MPLS has been popped exposing the payload packet.

To handle this, when a router (here it is router G) pops the final SR-MPLS label, it inserts an explicit null label [RFC3032] before encapsulating the packet in an MPLS-over-UDP tunnel toward router H and sending it out. That is, router G pops the top label, discovers it has reached the bottom of stack, pushes an explicit null label, and then encapsulates the MPLS packet in a UDP tunnel to router H.

3.2.2. Packet Forwarding without Penultimate Hop Popping

Figure 4 demonstrates the packet walk in the case where the label associated with each prefix-SID advertised by the owner of the prefix-SID is not a Penultimate Hop Popping (PHP) label (i.e., the the NP flag in OSPF or the P flag in ISIS associated with the prefix SID is set). Apart from the PHP function the roles of the routers is unchanged from Section 3.2.1.

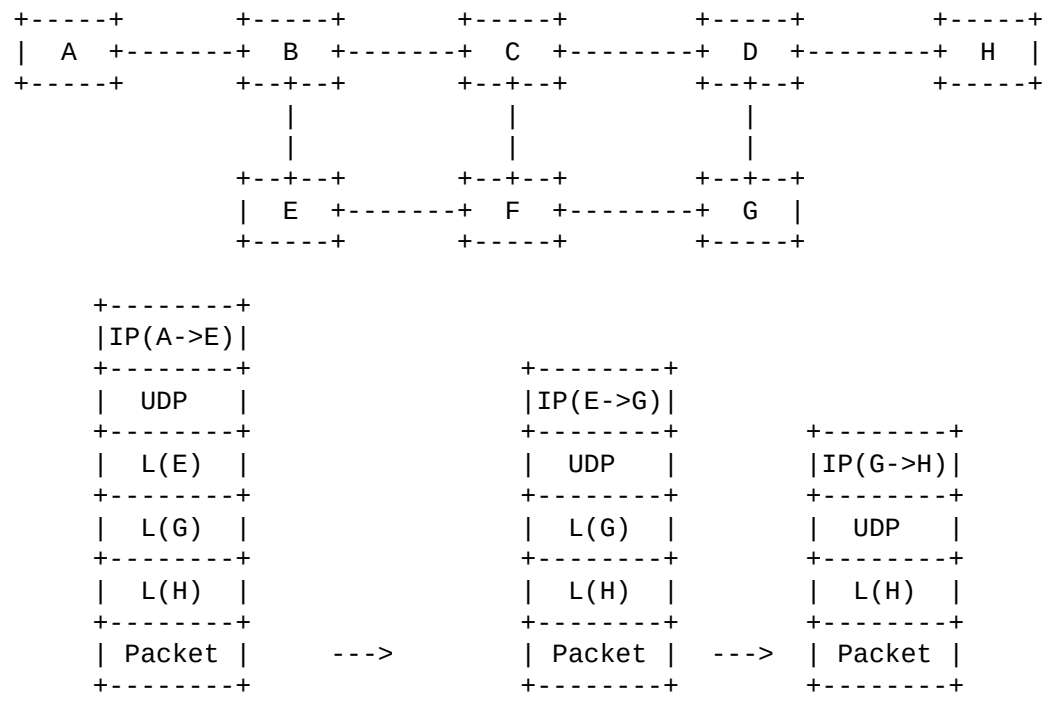

Figure 4: Packet Forwarding Example without PHP

As can be seen from the figure, the SR-MPLS label for each segment is left in place until the end of the segment where it is popped and the next instruction is processed.

3.2.3. Additional Forwarding Procedures

Non-MPLS Interfaces: Although the description in the previous two sections is based on the use of prefix-SIDs, tunneling SR-MPLS packets is useful when the top label of a received SR-MPLS packet indicates an adjacency-SID and the corresponding adjacent node to that adjacency-SID is not capable of MPLS forwarding but can still process SR-MPLS packets. In this scenario the top label would be replaced by an IP tunnel toward that adjacent node and then forwarded over the corresponding link indicated by the adjacency-SID.

When to use IP-based Tunnel: The description in the previous two sections is based on the assumption that MPLS-over-UDP tunnel is used when the nexthop towards the next segment is not MPLS-enabled. However, even in the case where the nexthop towards the next segment is MPLS-capable, an MPLS-over-UDP tunnel towards the

next segment could still be used instead due to local policies. For instance, in the example as described in Figure 4, assume F is now an SR-MPLS-capable transit node while all the other assumptions keep unchanged, since F is not identified by a SID in the stack and an MPLS-over-UDP tunnel is preferred to an MPLS LSP according to local policies, router E would replace the current top label with an MPLS-over-UDP tunnel toward router G and send it out.

IP Header Fields: When encapsulating an MPLS packet in UDP, the resulting packet is further encapsulated in IP for transmission. IPv4 or IPv6 may be used according to the capabilities of the network. The address fields are set as described in Section 2. The other IP header fields (such as DSCP code point, or IPv6 Flow Label) on each UDP-encapsulated segment SHOULD be configurable according to the operator's policy: they may be copied from the header of the incoming packet; they may be promoted from the header of the payload packet; they may be set according to instructions programmed to be associated with the SID; or they may be configured dependent on the outgoing interface and payload.

Entropy and ECMP: When encapsulating an MPLS packet with an IP tunnel header that is capable of encoding entropy (such as [RFC7510]), the corresponding entropy field (the source port in case UDP tunnel) MAY be filled with an entropy value that is generated by the encapsulator to uniquely identify a flow. However, what constitutes a flow is locally determined by the encapsulator. For instance, if the MPLS label stack contains at least one entropy label and the encapsulator is capable of reading that entropy label, the entropy label value could be directly copied to the source port of the UDP header. Otherwise, the encapsulator may have to perform a hash on the whole label stack or the five-tuple of the SR-MPLS payload if the payload is determined as an IP packet. To avoid re-performing the hash or hunting for the entropy label each time the packet is encapsulated in a UDP tunnel it MAY be desirable that the entropy value contained in the incoming packet (i.e., the UDP source port value) is retained when stripping the UDP header and is re-used as the entropy value of the outgoing packet.

4. IANA Considerations

This document makes no requests for IANA action.

5. Security Considerations

The security consideration of [RFC8354] and [RFC7510] apply. DTLS [RFC6347] SHOULD be used where security is needed on an MPLS-SR-over-UDP segment.

It is difficult for an attacker to pass a raw MPLS encoded packet into a network and operators have considerable experience at excluding such packets at the network boundaries.

It is easy for an ingress node to detect any attempt to smuggle an IP packet into the network since it would see that the UDP destination port was set to MPLS. SR packets not having a destination address terminating in the network would be transparently carried and would pose no security risk to the network under consideration.

Where control plane techniques are used (as described in Section 3), it is important that these protocols are adequately secured for the environment in which they are run.

6. Contributors

Ahmed Bashandy
Individual
Email: abashandy.ietf@gmail.com

Clarence Filsfils
Cisco
Email: cfilsfil@cisco.com

John Drake
Juniper
Email: jdrake@juniper.net

Shaowen Ma
Juniper
Email: mashao@juniper.net

Mach Chen
Huawei
Email: mach.chen@huawei.com

Hamid Assarpour
Broadcom
Email:hamid.assarpour@broadcom.com

Robert Raszuk
Bloomberg LP

Email: robert@raszuk.net

Uma Chunduri
Huawei
Email: uma.chunduri@gmail.com

Luis M. Contreras
Telefonica I+D
Email: luismiguel.contrerasmurillo@telefonica.com

Luay Jalil
Verizon
Email: luay.jalil@verizon.com

Gunter Van De Velde
Nokia
Email: gunter.van_de_velde@nokia.com

Tal Mizrahi
Marvell
Email: talmi@marvell.com

Jeff Tantsura
Individual
Email: jefftant@gmail.com

7. Acknowledgements

Thanks to Joel Halpern, Bruno Decraene, Loa Andersson, Ron Bonica, Eric Rosen, Jim Guichard, Gunter Van De Velde, Andy Malis, Robert Sparks, and Al Morton for their insightful comments on this draft.

8. References

8.1. Normative References

[I-D.ietf-spring-segment-routing-mpls]
Bashandy, A., Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with MPLS data plane", draft-ietf-spring-segment-routing-mpls-18 (work in progress), December 2018.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://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, <https://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, <https://www.rfc-editor.org/info/rfc3032>.

[RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", RFC 6347, DOI 10.17487/RFC6347, January 2012, <https://www.rfc-editor.org/info/rfc6347>.

[RFC7510] Xu, X., Sheth, N., Yong, L., Callon, R., and D. Black, "Encapsulating MPLS in UDP", RFC 7510, DOI 10.17487/RFC7510, April 2015, <https://www.rfc-editor.org/info/rfc7510>.

[RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, DOI 10.17487/RFC7684, November 2015, <https://www.rfc-editor.org/info/rfc7684>.

[RFC7794] Ginsberg, L., Ed., Decraene, B., Previdi, S., Xu, X., and U. Chunduri, "IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability", RFC 7794, DOI 10.17487/RFC7794, March 2016, <https://www.rfc-editor.org/info/rfc7794>.

[RFC7981] Ginsberg, L., Previdi, S., and M. Chen, "IS-IS Extensions for Advertising Router Information", RFC 7981, DOI 10.17487/RFC7981, October 2016, <https://www.rfc-editor.org/info/rfc7981>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

8.2. Informative References

[I-D.ietf-6man-segment-routing-header]
Filsfils, C., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-16 (work in progress), February 2019.

[I-D.ietf-isis-encapsulation-cap]
Xu, X., Decraene, B., Raszuk, R., Chunduri, U., Contreras, L., and L. Jalil, "Advertising Tunnelling Capability in IS-IS", draft-ietf-isis-encapsulation-cap-01 (work in progress), April 2017.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-22 (work in progress), December 2018.

[I-D.ietf-mpls-spring-entropy-label]
Kini, S., Kompella, K., Sivabalan, S., Litkowski, S., Shakir, R., and J. Tantsura, "Entropy label for SPRING tunnels", draft-ietf-mpls-spring-entropy-label-12 (work in progress), July 2018.

[I-D.ietf-ospf-encapsulation-cap]
Xu, X., Decraene, B., Raszuk, R., Contreras, L., and L. Jalil, "The Tunnel Encapsulations OSPF Router Information", draft-ietf-ospf-encapsulation-cap-09 (work in progress), October 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-27 (work in progress), December 2018.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[RFC8354] Brzozowski, J., Leddy, J., Filsfils, C., Maglione, R., Ed., and M. Townsley, "Use Cases for IPv6 Source Packet Routing in Networking (SPRING)", RFC 8354, DOI 10.17487/RFC8354, March 2018, <https://www.rfc-editor.org/info/rfc8354>.

Authors' Addresses

Xiaohu Xu
Alibaba, Inc

Email: xiaohu.xxh@alibaba-inc.com

Stewart Bryant
Huawei

Email: stewart.bryant@gmail.com

Adrian Farrel
Old Dog Consulting

Email: adrian@olddog.co.uk

Syed Hassan
Cisco

Email: shassan@cisco.com

Wim Henderickx
Nokia

Email: wim.henderickx@nokia.com

Zhenbin Li
Huawei

Email: lizhenbin@huawei.com