

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 1, 2018

X. Xu
Huawei
A. Bashandy
Cisco
H. Assarpour
Broadcom
S. Ma
Juniper
W. Henderickx
Nokia
J. Tantsura
Individual
September 28, 2017

Unified Source Routing Instructions using MPLS Label Stack
draft-xu-mpls-unified-source-routing-instruction-04

Abstract

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

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 April 1, 2018.

Copyright Notice

Copyright (c) 2017 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	2
1.1. Requirements Language	3
2. Terminology	3
3. Use Cases	3
4. Packet Forwarding Procedures	4
4.1. Forwarding Entry Construction	5
4.2. Packet Forwarding Procedures	6
5. Contributors	9
6. Acknowledgements	10
7. IANA Considerations	10
8. Security Considerations	10
9. References	10
9.1. Normative References	10
9.2. Informative References	10
Authors' Addresses	13

1. Introduction

MPLS Segment Routing (SR-MPLS in short) [I-D.ietf-spring-segment-routing-mpls] is an MPLS data plane-based source routing paradigm in which a sender of a packet is allowed to partially or completely specify the route the packet takes through the network by imposing stacked MPLS labels to the packet. SR-MPLS could be leveraged to realize a unified source routing mechanism across MPLS, IPv4 and IPv6 data planes by using an MPLS label stack as a unified source routing instruction set while preserving 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 the underlay is IPv4, IPv6 or MPLS.

Although the source routing instructions are encoded as MPLS labels, this is a hardware convenience rather than an indication that the whole MPLS protocol stack and in particular the MPLS control protocols need to be deployed. Note that the complexity associated with the whole MPLS protocol stack is largely due to the complex control plane protocols.

Section 3 describes various use cases for the unified source routing instruction mechanism and Section 4 describes a typical application scenario and how the packet forwarding happens.

1.1. 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].

2. Terminology

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

3. Use Cases

The unified source routing mechanism across IPv4, IPv6 and MPLS is useful at least in the following use cases:

- o Incremental deployment of the SR-MPLS technology [I-D.xu-mpls-spring-islands-connection-over-ip]. Since there is no need to run any other label distribution protocol (e.g., LDP, see [I-D.ietf-spring-segment-routing-ldp-interop] for more details.) on those non-SR-MPLS routers for incremental deployment purposes, the network provisioning is greatly simplified, which is one of the major claimed benefits of the SR-MPLS technology (i.e., running a single protocol).
- o Overcome the load-balancing dilemma encountered by SR-MPLS. In fact, this unified source routing mechanism is even useful in a fully upgraded SR-MPLS network since the load-balancing dilemma encountered by SR-MPLS [I-D.ietf-mpls-spring-entropy-label] due to the maximum Readable Label-stack Depth (RLD) hardware limitation [I-D.ietf-ospf-mpls-elc] [I-D.ietf-isis-mpls-elc] [I-D.ietf-idr-bgp-ls-segment-routing-rld] and the Maximum SID Depth (MSD) hardware limitation [I-D.ietf-ospf-segment-routing-msd] [I-D.ietf-isis-segment-routing-msd] [I-D.ietf-idr-bgp-ls-segment-routing-msd] by using the MPLS-in-UDP

encapsulation [RFC7510] where the source port of the UDP tunnel header is used as an entropy field.

- o A poor man's light-weight alternative to SRv6 [I-D.ietf-6man-segment-routing-header]. At least, it could be deployed as an interim until full featured SRv6 is available on more platforms. Since the Source Routing Header (SRH) [I-D.ietf-6man-segment-routing-header] consisting of an ordered list of 128-bit long IPv6 addresses is now replaced by an ordered list of 32-bit long label entries (i.e., label stack), the encapsulation overhead and forwarding performance issues associated with SRv6 are eliminated.
- o A new IPv4 source routing mechanism which has overcome the security vulnerability issues associated with the traditional IPv4 source routing mechanism.
- o Traffic Engineering scenarios where only a few routers (e.g., the entry and exit nodes of each plane in the dual-plane network case or the egress node in the Egress Peer Engineering (EPE) case) are specified as segments of explicit paths. In this way, only a few routers are required to support the SR-MPLS capability while all the other routers just need to support IP forwarding capability, which would significantly reduce the deployment cost of the SR-MPLS technology.
- o MPLS-based Service Function Chaining (SFC) [I-D.xu-mpls-service-chaining]. Based on the unified source routing mechanism as described in this document, only SFC-related nodes including Service Function Forwarders (SFF), Service Functions (SF) and classifiers are required to recognize the SFC encapsulation header in the MPLS label stack form, while the intermediate routers just need to support vanilla IP forwarding (either IPv4 or IPv6). In other words, it undoubtedly complies with the transport-independence requirement for the SFC encapsulation header as listed in the SFC architecture document [RFC7665].

4. Packet Forwarding Procedures

The primary objective of this document is to describe how SR-MPLS capable routers and IP-only routers can seamlessly co-exist and interoperate. This section describes the forwarding information base (FIB) entry and the forwarding behavior that allow the deployment of SR-MPLS when some routers are IPv4 only or IPv6 only. Note that OSPF or ISIS is assumed to be enabled in the following examples as described in Section 4.1 and 4.2, in fact, it's no doubt that BGP could be used as a replacement.

4.1. 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-SID are IPv4-only or IPv6-only routers. Consider the router "A" receiving a labeled packet whose top label L(E) corresponds to the prefix-SID is "SID(E)" of prefix "P(E)" advertised by the router "E". Suppose the *i*th next-hop router "NHi" along the shortest path from the router "A" towards the prefix-SID "SID(E)" is not SR-MPLS capable. That is both routers "A" and "E" are SR-MPLS capable but the next hop "NHi" along the shortest path from "A" to "E". The following applies:

- o It is assumed that the router "E" advertises the SR-Capabilities sub-TLV as described in and [I-D.ietf-ospf-segment-routing-extensions], which includes the SRGB because router "E" is SR-MPLS capable.
- o The owning router "E" MUST advertise the encapsulation endpoint and the tunnel type using [I-D.ietf-isis-encapsulation-cap] and/or [I-D.ietf-ospf-encapsulation-cap] .
- o If "A" and "E" are in different areas/levels, then
 - * The OSPF Tunnel Encapsulation TLV [I-D.ietf-ospf-encapsulation-cap] and/or the ISIS Tunnel Encapsulation sub-TLV [I-D.ietf-isis-encapsulation-cap] are flooded domain-wide.
 - * The OSPF SID/label range TLV [I-D.ietf-ospf-segment-routing-extensions] and the ISIS SR-Capabilities Sub-TLV [I-D.ietf-isis-segment-routing-extensions] are advertised domain-wide. This way router "A" knows the characteristics of the owning router "E".
 - * When the owning router "E" is running ISIS and advertises the prefix "P(E) ", the router "E" uses the extended reachability TLV (TLVs 135, 235, 236, 237) and associates the IPv4/IPv6 and/or IPv4/IPv6 source router ID sub-TLV(s) [RFC7794].
 - * When the owning router "E" is running OSPF and advertises the prefix "P(E)", the router "E" uses the OSPFv2 Extended Prefix Opaque LSA [RFC7684] and sets the flooding scope to AS-wide.
 - * When the owning 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 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" .

- o Router "A" programs the FIB entry corresponding to the "SID(E)" as follows:
 - * If NP (OSPF) or P (ISIS) flag is clear,
 - *
 - + pop the outer label.
 - * If NP (OSPF) or P (ISIS) is set,
 - *
 - + the outer label is SID(E) plus the lower bound of the SRGB of "E".
 - * Encapsulate the packet according to the encapsulation advertised in [I-D.ietf-isis-encapsulation-cap] or [I-D.ietf-ospf-encapsulation-cap].
 - * Send the packet towards the next hop "NH_i".

4.2. Packet Forwarding Procedures

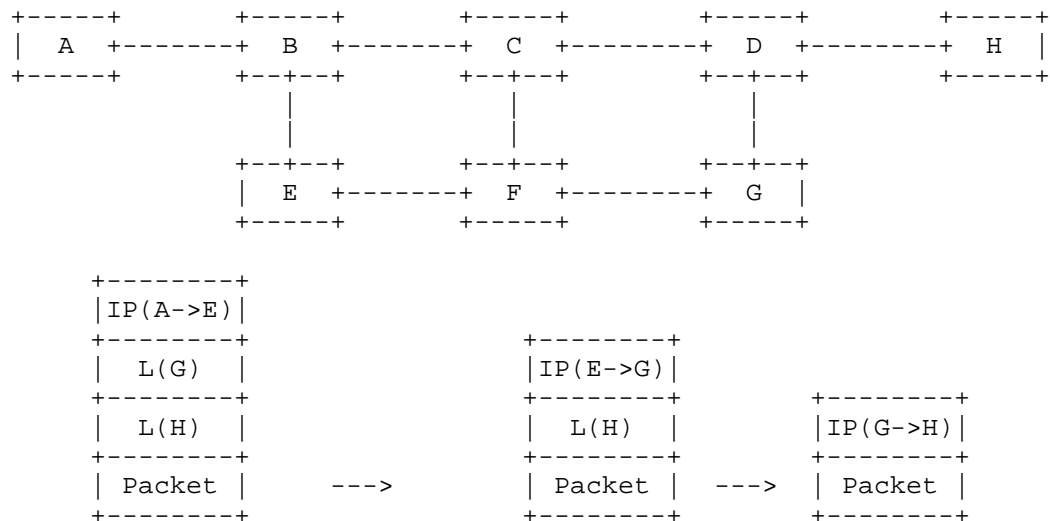


Figure 1

As shown in Figure 1, Assume Router A, E, G and H are SR-MPLS-capable routers while the remaining are only capable of forwarding IP packets. Router A, E, G and H advertise their Segment Routing related information via IS-IS or OSPF. Now assume router A wants to send a given IP or MPLS packet via an explicit path of {E->G->H}, router A would impose an MPLS label stack corresponding to that explicit path on the received IP packet. Since there is no Label Switching Path (LSP) towards router E, router A would replace the top label indicating router E with an IP-based tunnel for MPLS (e.g., MPLS-over-UDP [RFC7510]) towards router E and then send it out. In other words, router A would pop the top label and then encapsulate the MPLS packet with an IP-based tunnel towards router E. When the IP-encapsulated MPLS packet arrives at router E, router E would strip the IP-based tunnel header and then process the decapsulated MPLS packet accordingly. Since there is no LSP towards router G which is indicated by the current top label of the decapsulated MPLS packet, router E would replace the current top label with an IP-based tunnel towards router G and send it out. When the packet arrives at router G, router G would strip the IP-based tunnel header and then process the decapsulated MPLS packet. Since there is no LSP towards router H, router G would replace the current top label with an IP-based tunnel towards router H. Now the packet encapsulated with the IP-based tunnel towards router H is exactly the original packet that router A had intended to send towards router H. If the packet is an MPLS packet, router G could use any IP-based tunnel for MPLS (e.g., MPLS-over-UDP [RFC7510]). If the packet is an IP packet, router G could use any IP tunnel for IP (e.g., IP-in-UDP [I-D.xu-intarea-ip-in-udp]). That original IP or MPLS packet would be forwarded towards router H via an IP-based tunnel. When the encapsulated packet arrives at router H, router H would decapsulate it into the original packet and then process it accordingly.

Note that in the above description, it's assumed that the label associated with each prefix-SID advertised by the owner of the prefix-SID is a Penultimate Hop Popping (PHP) label (e.g., the NP-flag [I-D.ietf-ospf-segment-routing-extensions] associated with the corresponding prefix SID is not set).

Figure 2 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 (e.g., the NP-flag [I-D.ietf-ospf-segment-routing-extensions] associated with the corresponding prefix SID is set).

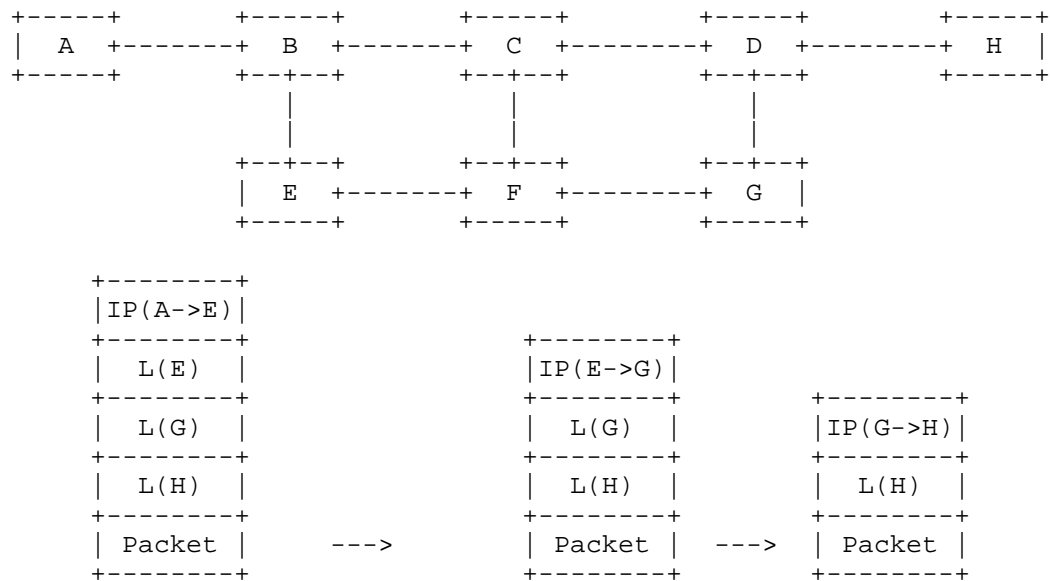


Figure 2

Although the above description is based on the use of prefix-SIDs, the unified source routing instruction approach is actually applicable to the use of adj-SIDs as well. For instance, when the top label of a received MPLS packet indicates an given adj-SID and the corresponding adjacent node to that adj-SID is not MPLS-capable, the top label would be replaced by an IP-based tunnel towards that adjacent node and then forwarded over the corresponding link indicated by that adj-SID.

When encapsulating an MPLS packet with an IP-based tunnel header (e.g., a UDP header as per [RFC7510]), the corresponding entropy field (i.e., the source port in the MPLS-in-UDP case) should 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 entropy field (e.g., 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 MPLS payload if the payload is determined as an IP packet. To avoid re-performing hash on the whole packet when re-encapsulating the packet with an IP-based tunnel header (e.g., a UDP tunnel header), especially when the encapsulator could not obtain at least one entropy label due to some reasons (e.g., 1) there is no EL at all in the label stack; 2) the encapsulator couldn't recognize the ELI; 3) the encapsulator could

not read the EL due to the RLD limit), it's RECOMMENDED that the entropy value contained in the packet (e.g., the UDP source port value) is kept when stripping the IP-based tunnel header (e.g., the UDP tunnel header). As such, the entropy value could be directly copied to the entropy field (e.g., the source port of the UDP tunnel header) when re-encapsulating the packet with an IP-based tunnel header (e.g., a UDP tunnel header). As such, the load-balancing dilemma encountered by SR-MPLS as described in [I-D.ietf-mpls-spring-entropy-label] due to the maximum Readable Label-stack Depth (RLD) hardware limitation [I-D.ietf-ospf-mpls-elc] [I-D.ietf-isis-mpls-elc] and the Maximum SID Depth (MSD) hardware limitation [I-D.ietf-ospf-segment-routing-msd] [I-D.ietf-isis-segment-routing-msd] is gone. That's the reason why this unified source routing mechanism is even useful in a fully upgraded SR-MPLS network environment.

5. Contributors

Clarence Filsfils
Cisco
Email: cfilsfil@cisco.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

6. Acknowledgements

Thanks Joel Halpern, Bruno Decraene, Loa Andersson and Stewart Bryant for their insightful comments on this document.

7. IANA Considerations

No IANA action is required.

8. Security Considerations

TBD.

9. References

9.1. 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, <<https://www.rfc-editor.org/info/rfc2119>>.

9.2. Informative References

[I-D.ietf-6man-segment-routing-header]
Previdi, S., Filsfils, C., Raza, K., Leddy, J., Field, B., daniel.voyer@bell.ca, d., daniel.bernier@bell.ca, d., Matsushima, S., Leung, I., Linkova, J., Aries, E., Kosugi, T., Vyncke, E., Lebrun, D., Steinberg, D., and R. Raszuk, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-07 (work in progress), July 2017.

[I-D.ietf-idr-bgp-ls-segment-routing-msd]
Tantsura, J., Chunduri, U., Mirsky, G., and S. Sivabalan, "Signaling Maximum SID Depth using Border Gateway Protocol Link-State", draft-ietf-idr-bgp-ls-segment-routing-msd-00 (work in progress), July 2017.

[I-D.ietf-idr-bgp-ls-segment-routing-rld]
Velde, G., Henderickx, W., Bocci, M., and K. Patel, "Signalling ERLD using BGP-LS", draft-ietf-idr-bgp-ls-segment-routing-rld-00 (work in progress), July 2017.

[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-mpls-elc]

Xu, X., Kini, S., Sivabalan, S., Filsfils, C., and S. Litkowski, "Signaling Entropy Label Capability Using IS-IS", draft-ietf-isis-mpls-elc-02 (work in progress), October 2016.

[I-D.ietf-isis-segment-routing-extensions]

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

[I-D.ietf-isis-segment-routing-msd]

Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling MSD (Maximum SID Depth) using IS-IS", draft-ietf-isis-segment-routing-msd-04 (work in progress), June 2017.

[I-D.ietf-mpls-spring-entropy-label]

Kini, S., Kompella, K., Sivabalan, S., Litkowski, S., Shakir, R., and j. jeffrant@gmail.com, "Entropy label for SPRING tunnels", draft-ietf-mpls-spring-entropy-label-06 (work in progress), May 2017.

[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-08 (work in progress), September 2017.

[I-D.ietf-ospf-mpls-elc]

Xu, X., Kini, S., Sivabalan, S., Filsfils, C., and S. Litkowski, "Signaling Entropy Label Capability Using OSPF", draft-ietf-ospf-mpls-elc-04 (work in progress), November 2016.

[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-19 (work in progress), August 2017.

[I-D.ietf-ospf-segment-routing-msd]

Tantsura, J., Chunduri, U., Aldrin, S., and P. Psenak, "Signaling MSD (Maximum SID Depth) using OSPF", draft-ietf-ospf-segment-routing-msd-05 (work in progress), June 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.
- [I-D.ietf-spring-segment-routing-mpls]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with MPLS data plane", draft-ietf-spring-segment-routing-mpls-10 (work in progress), June 2017.
- [I-D.xu-intarea-ip-in-udp]
Xu, X., Lee, Y., and F. Yongbing, "Encapsulating IP in UDP", draft-xu-intarea-ip-in-udp-04 (work in progress), December 2016.
- [I-D.xu-mpls-service-chaining]
Xu, X., Bryant, S., Assarpour, H., Shah, H., Contreras, L., daniel.bernier@bell.ca, d., jefftant@gmail.com, j., Ma, S., and M. Vigoureux, "Service Chaining using Unified Source Routing Instructions", draft-xu-mpls-service-chaining-03 (work in progress), June 2017.
- [I-D.xu-mpls-spring-islands-connection-over-ip]
Xu, X., Raszuk, R., Chunduri, U., Contreras, L., and L. Jalil, "Connecting MPLS-SPRING Islands over IP Networks", draft-xu-mpls-spring-islands-connection-over-ip-00 (work in progress), October 2016.
- [RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, DOI 10.17487/RFC2784, March 2000, <<https://www.rfc-editor.org/info/rfc2784>>.
- [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>>.
- [RFC4817] Townsley, M., Pignataro, C., Wainner, S., Seely, T., and J. Young, "Encapsulation of MPLS over Layer 2 Tunneling Protocol Version 3", RFC 4817, DOI 10.17487/RFC4817, March 2007, <<https://www.rfc-editor.org/info/rfc4817>>.

- [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>>.
- [RFC7665] Halpern, J., Ed. and C. Pignataro, Ed., "Service Function Chaining (SFC) Architecture", RFC 7665, DOI 10.17487/RFC7665, October 2015, <<https://www.rfc-editor.org/info/rfc7665>>.
- [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>>.

Authors' Addresses

Xiaohu Xu
Huawei

Email: xuxiaohu@huawei.com

Ahmed Bashandy
Cisco

Email: bashandy@cisco.com

Hamid Assarpour
Broadcom

Email: hamid.assarpour@broadcom.com

Shaowen Ma
Juniper

Email: mashao@juniper.net

Wim Henderickx
Nokia

Email: wim.henderickx@nokia.com

Jeff Tantsura
Individual

Email: jefftant@gmail.com