

mboned
Internet-Draft
Intended status: Informational
Expires: 25 October 2025

Z. Zhang
Juniper Networks
I. Wijnands
Arrcus
H. Bidgoli
Nokia
Y. Liu
China Mobile
23 April 2025

Non-source-routed Multicast in SR Networks
draft-ietf-mboned-non-source-routed-sr-mcast-02

Abstract

With Segment Routing (SR) architecture, a unicast flow can be source-routed through an SR network following an explicit path specified in the packet, w/o the need for per-flow state in the network. As a result, the otherwise needed protocols to signal the per-flow unicast state can also be removed from the network. In the case of multicast, traffic can be either source-routed or non-source-routed, and this document discusses non-sourced-routed options for multicast in an SR network with either MPLS or IPv6/SRv6 data plane.

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 25 October 2025.

Copyright Notice

Copyright (c) 2025 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
2. Traditional Multicast Technologies	3
3. Bit Index Explicit Replication	4
4. Multicast Trees Set up by Other Means	4
5. SR Specific Solutions	5
6. IPv6 Considerations	5
7. Summary	6
8. Security Considerations	7
9. Acknowledgements	7
10. Informative References	7
Authors' Addresses	10

1. Introduction

With Segment Routing (SR) architecture [RFC8402], a unicast flow can be source-routed through an SR network following an explicit path specified in the packet, without the need for per-flow state in the network. As a result, the otherwise needed protocols to signal the per-flow unicast state can also be removed from the network. This "no per-flow state" is the number one principle of SR, and the optional use of controllers is another characteristic of SR.

This document summarizes options for non-source-routed multicast in an SR network, including traditional multicast technologies, BIER, and various SR specific solutions. The pros and cons of each solution are listed for considerations by operators and vendors, with regard to the principle and characteristic of SR as mentioned above.

Source-routed multicast can be done with BIER-TE [RFC9262], and, in the case of SRv6 network it has been discussed in IETF [<https://datatracker.ietf.org/doc/bofreg-liu-multicast-source-routing-over-ipv6msr6/>] and some work is ongoing. However, that is outside the scope of this document.

As discussed in [I-D.zzhang-pim-multicast-scaling-considerations], end-to-end multicast is best via IP multicast trees, though they could be transported over some (underlay) tunnels. In that regard,

when we talk about multicast in SR networks, we refer to the multicast in the underlay SR network - whether it is for underlay tunnels transporting overlay multicast (e.g., [RFC6513]), or for end-to-end IP multicast directly through (vs. over) the SR network.

2. Traditional Multicast Technologies

Traditional multicast technologies include PIM [RFC7761], RSVP-TE P2MP [RFC4875], and mLDP [RFC6388]. They all require per-tree state on nodes on the tree, and the corresponding protocols to signal and maintain the state. An incoming packet's IP header or MPLS label is looked up, and the packet is forwarded according to the matched state.

While SR allows simplification of state, protocols and centralized SDN-control, the traditional methods of delivering multicast traffic run contrary to those SR goals.

An alternative is Ingress Replication (IR) - an upstream node of a multicast packet tunnels individual copies to some downstream nodes across some intermediate nodes. In an SR network, the unicast tunnels used for IR could be traditional IP/MPLS tunnels or could be SR tunnels (referred to as SR policies), which could be MPLS/SRv6-based.

While with IR there is no per-tree state on those intermediate nodes, multiple copies of the same packet may be sent over the same link. If efficient replication is required, PIM/mLDP/RSVP-TE P2MP can still be used for multicast even in an SR network. Deploying SR in the network does not mandate changing the solution that is in place for Multicast.

While mLDP uses the same LDP Session (and protocol packet encoding) as used by unicast and there is code sharing between LDP and mLDP with regards to neighbor discovery, session setup and Label allocation management, the protocol logic for setting up mLDP tunnel is completely different. It is understood that there is a desire to remove LDP from the network when SR is deployed, but there is really no problem to continue to run mLDP for Multicast. [RFC7473] documents a solution where LDP can be run in mLDP-Only mode by simply not exchanging any Unicast bindings.

Similarly, RSVP-TE P2MP tunnel setup is orthogonal from P2P tunnel setup as well. It can be used for multicast even in an SR network if bandwidth reservation and/or per-tunnel explicit path are required.

3. Bit Index Explicit Replication

Bit Index Explicit Replication (BIER) [RFC8279] is a new multicast technology that achieves efficient replication as with PIM tree or mLDP/RSVP-TE P2MP tunnels but without requiring per-tree/tunnel state in the transit nodes as with IR.

Even though BIER is designed and developed independently of SR, it is a perfect fit for an SR network because it shares with SR the number one principle of "no per-flow state". A BIER domain can be used as provider/underlay tunnels for MVPN/EVPN-BUM (just like traditional PIM tree or mLDP/RSVP P2MP tunnels), or can be part of end-to-end PIM trees [I-D.ietf-bier-pim-signaling] (similar to mLDP inband signaling [RFC6826]).

However, BIER uses a new forwarding plane that cannot be implemented on many deployed routers. On the other hand, the entry point barrier is decreasing and BIER is becoming more realistic with the following developments:

- * Several major ASIC makers have BIER support on their ASIC offering.
- * Several Major router vendors have BIER hardware capability across their access/aggregation/edge/core platforms.
- * Some major router vendors have had or will soon have interoperable production support of BIER on certain platforms. Roadmaps for BIER on more platforms also exist.
- * There have been very good discussions and progress in BIER WG for methods of brown field deployments ([I-D.przygienda-bier-migration-options] [I-D.ietf-bier-tether] [I-D.ietf-bier-multicast-as-a-service] [I-D.ietf-bier-php])
- * More and more customers now have concrete plans and requirements for BIER deployment in their networks

4. Multicast Trees Set up by Other Means

Some operators may accept that they need to maintain per-tree state in their SR network for efficient multicast, and their applications do not require too much per-tree multicast state. However, since unicast no longer needs LDP or RSVP-TE protocol in an SR network, they really want to remove PIM/LDP/RSVP-TE protocol from their networks, and make use of controllers. In that case, the multicast trees could be set up statically from management plane (e.g., netconf), or dynamically via BGP/PCEP [I-D.ietf-bess-bgp-multicast]

[I-D.ietf-bess-bgp-multicast-controller]
[I-D.ietf-idr-sr-p2mp-policy] [I-D.ietf-pce-sr-p2mp-policy].

5. SR Specific Solutions

There are two multicast solutions that are specifically tied to SR technologies - "SR-P2MP" and "Spray".

1. SR-P2MP refers to trees setup not by mLDP or RSVP-TE. It could be statically configured, or instantiated from a controller, as described in [I-D.ietf-pim-sr-p2mp-policy]. It uses Replication Segments specified in [RFC9524], which describe the replication states on the tree nodes.
2. Spray is Ingress Replication with SRv6. Instead of regular IP/MPLS tunneling, an SRv6 header encodes the "tunnel end" and the original destination multicast address. When the packet arrives at the tunnel end, the original multicast address is restored from the SRv6 header.

The above methods are really no different from existing technologies. Spray is just IR with a different encapsulation (e.g. SRv6). While SR-P2MP does not use existing tree signaling protocol, it still requires per-tree forwarding state on the tree nodes (roots, leaves and branching points), and it is identical to mLDP/RSVP P2MP tunnels in the MPLS data plane on those tree nodes. In other words, SR-P2MP does not follow the number one principle of SR after all.

mLDP/PIM trees are typically hop-by-hop - the tree nodes are directly connected, though tunnels can be used between mLDP, PIM and SR-P2MP tree nodes to avoid tree states on the non-branching nodes between tree nodes. This is important in some deployment scenarios (e.g., mobile user plane transport) where the replication points are scattered with many non-replication points among them.

SR-P2MP can use both MPLS and SRv6 data plane. The next section discusses some SRv6 aspects of SR-P2MP.

6. IPv6 Considerations

Some operators run IPv6/SRv6 networks without using MPLS at all. As a result, Ingress Replication or SR-P2MP with MPLS data plane is not an option and IPv6 based forwarding must be used. There are two options in this case - traditional IPv6 multicast (signaled by either PIM or BGP [I-D.ietf-bess-bgp-multicast] [I-D.ietf-bess-bgp-multicast-controller]) and SRv6-P2MP (SR-P2MP with SRv6 data plane).

SR-P2MP with MPLS data plane (SRmpls-P2MP) is identical to mLDP/RSVP-P2MP in the data plane of the tree nodes, in that incoming multicast traffic carries a tree-identifying label and is replicated with outgoing label stacks. The label stack includes a label that identifies the tree to the downstream node, and optional labels that leads the traffic to the downstream node.

SRv6-P2MP is very similar to SRmpls-P2MP in the data plane. Multicast traffic has an IPv6 header and the destination address has the LOC:FUNCT:ARGS encoding format for SRv6 [RFC8986]. The FUNCT bits correspond to the tree-identifying label in SRmpls-P2MP, and the LOC bits get the traffic from an upstream node to a downstream node (which could be directly or indirectly connected).

SRv6-P2MP allows native tunneling of multicast traffic between indirectly connected upstream and downstream nodes, but at the cost of changing destination address at each tree node, which requires either programmable ASIC or new generation of non-programmable ASIC. It also requires a new control plane that involves controller-based calculation and signaling of the tree state.

On the other hand, traditional IPv6 multicast is mature in both control plane and data plane. PIM has been widely deployed for many years without platform restrictions, and the PIM-SSM protocol is very straightforward without the complexity of PIM-ASM.

Even if an operator wants to use controller based calculation and signaling, or does not want to run PIM protocol, traditional multicast can still be used with BGP/controller-based signaling as specified in [I-D.ietf-bess-bgp-multicast] and [I-D.ietf-bess-bgp-multicast-controller].

7. Summary

There is no one-for-all option when it comes to multicast in an SR network. Depending on specific deployment scenarios and requirements, the following could be considered in order when source-routing is not used:

- * If most of nodes can support it, BIER is the best choice, because it removes state from the network and simplifies the control-plane (like SR does).
- * If efficient multicast replication is required, mLDP/RSVP-TE/PIM can still be used for multicast purpose if that is acceptable.

- * If the replication points are scattered with many non-replication points between them, or if there is a strong desire to create multicast forwarding state without relying on PIM/mLDP or RSVP-TE, one can use static configuration (e.g., netconf based provisioning) or controller-based tree calculation/signaling (SR-P2MP, BGP-signaled IP multicast or mLDP).
- * Use Ingress Replication with SR unicast tunnels.

In case of MPLS based P2MP, a Binding SID could be optionally used to direct traffic to the root of the tunnel.

8. Security Considerations

This informational document does not introduce new security risks.

9. Acknowledgements

The authors thank Eric Rosen, Tony Przygienda, and John Drake for their reviews, comments and suggestions.

10. Informative References

[I-D.ietf-bess-bgp-multicast]

Zhang, Z. J., Giuliano, L., Patel, K., Wijnands, I., Mishra, M. P., and A. Gulko, "BGP Based Multicast", Work in Progress, Internet-Draft, draft-ietf-bess-bgp-multicast-08, 3 June 2024, <<https://datatracker.ietf.org/doc/html/draft-ietf-bess-bgp-multicast-08>>.

[I-D.ietf-bess-bgp-multicast-controller]

Zhang, Z. J., Raszuk, R., Pacella, D., and A. Gulko, "Controller-based BGP Multicast Signaling", Work in Progress, Internet-Draft, draft-ietf-bess-bgp-multicast-controller-16, 28 February 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-bess-bgp-multicast-controller-16>>.

[I-D.ietf-bier-multicast-as-a-service]

Zhang, Z. J., Rosen, E. C., Awduche, D. O., Shepherd, G., Zhang, Z., and G. S. Mishra, "Multicast/BIER As A Service", Work in Progress, Internet-Draft, draft-ietf-bier-multicast-as-a-service-03, 6 February 2024, <<https://datatracker.ietf.org/doc/html/draft-ietf-bier-multicast-as-a-service-03>>.

[I-D.ietf-bier-php]

Zhang, Z. J., "BIER Penultimate Hop Popping", Work in Progress, Internet-Draft, draft-ietf-bier-php-16, 4 December 2024, <<https://datatracker.ietf.org/doc/html/draft-ietf-bier-php-16>>.

[I-D.ietf-bier-pim-signaling]

Bidgoli, H., Xu, F., Kotalwar, J., Wijnands, I., Mishra, M. P., and Z. J. Zhang, "PIM Signaling Through BIER Core", Work in Progress, Internet-Draft, draft-ietf-bier-pim-signaling-13, 3 March 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-bier-pim-signaling-13>>.

[I-D.ietf-bier-tether]

Zhang, Z. J., Warnke, N., Wijnands, I., and D. O. Awduche, "Tethering A BIER Router To A BIER Incapable Router", Work in Progress, Internet-Draft, draft-ietf-bier-tether-08, 3 December 2024, <<https://datatracker.ietf.org/doc/html/draft-ietf-bier-tether-08>>.

[I-D.ietf-idr-sr-p2mp-policy]

Bidgoli, H., Voyer, D., Stone, A., Parekh, R., Krier, S., and S. Agrawal, "Advertising p2mp policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-sr-p2mp-policy-00, 27 May 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-p2mp-policy-00>>.

[I-D.ietf-pce-sr-p2mp-policy]

Bidgoli, H., Voyer, D., Budhiraja, A., Parekh, R., and S. Sivabalan, "PCEP extensions for P2MP SR Policy", Work in Progress, Internet-Draft, draft-ietf-pce-sr-p2mp-policy-11, 19 February 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-sr-p2mp-policy-11>>.

[I-D.ietf-pim-sr-p2mp-policy]

Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., Zhang, Z. J., and M. P. Mishra, "Segment Routing Point-to-Multipoint Policy", Work in Progress, Internet-Draft, draft-ietf-pim-sr-p2mp-policy-11, 2 February 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-pim-sr-p2mp-policy-11>>.

[I-D.przygienda-bier-migration-options]

Przygienda, T. and Z. J. Zhang, "BIER Migration Frameworks", Work in Progress, Internet-Draft, draft-

przygienda-bier-migration-options-00, 22 June 2018,
<<https://datatracker.ietf.org/doc/html/draft-przygienda-bier-migration-options-00>>.

- [I-D.zzhang-pim-multicast-scaling-considerations]
Zhang, Z. J., Parekh, R., Bidgoli, H., and Z. Zhang,
"Multicast Scaling Considerations", Work in Progress,
Internet-Draft, draft-zzhang-pim-multicast-scaling-
considerations-00, 24 October 2022,
<<https://datatracker.ietf.org/doc/html/draft-zzhang-pim-multicast-scaling-considerations-00>>.
- [RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S.
Yasukawa, Ed., "Extensions to Resource Reservation
Protocol - Traffic Engineering (RSVP-TE) for Point-to-
Multipoint TE Label Switched Paths (LSPs)", RFC 4875,
DOI 10.17487/RFC4875, May 2007,
<<https://www.rfc-editor.org/info/rfc4875>>.
- [RFC6388] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B.
Thomas, "Label Distribution Protocol Extensions for Point-
to-Multipoint and Multipoint-to-Multipoint Label Switched
Paths", RFC 6388, DOI 10.17487/RFC6388, November 2011,
<<https://www.rfc-editor.org/info/rfc6388>>.
- [RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/
BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February
2012, <<https://www.rfc-editor.org/info/rfc6513>>.
- [RFC6826] Wijnands, IJ., Ed., Eckert, T., Leymann, N., and M.
Napierala, "Multipoint LDP In-Band Signaling for Point-to-
Multipoint and Multipoint-to-Multipoint Label Switched
Paths", RFC 6826, DOI 10.17487/RFC6826, January 2013,
<<https://www.rfc-editor.org/info/rfc6826>>.
- [RFC7473] Raza, K. and S. Boutros, "Controlling State Advertisements
of Non-negotiated LDP Applications", RFC 7473,
DOI 10.17487/RFC7473, March 2015,
<<https://www.rfc-editor.org/info/rfc7473>>.
- [RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I.,
Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent
Multicast - Sparse Mode (PIM-SM): Protocol Specification
(Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March
2016, <<https://www.rfc-editor.org/info/rfc7761>>.

- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [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>>.
- [RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <<https://www.rfc-editor.org/info/rfc8986>>.
- [RFC9262] Eckert, T., Ed., Menth, M., and G. Cauchie, "Tree Engineering for Bit Index Explicit Replication (BIER-TE)", RFC 9262, DOI 10.17487/RFC9262, October 2022, <<https://www.rfc-editor.org/info/rfc9262>>.
- [RFC9524] Voyer, D., Ed., Filsfils, C., Parekh, R., Bidgoli, H., and Z. Zhang, "Segment Routing Replication for Multipoint Service Delivery", RFC 9524, DOI 10.17487/RFC9524, February 2024, <<https://www.rfc-editor.org/info/rfc9524>>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks
Email: zzhang@juniper.net

IJsbrand Wijnands
Arcus
Email: ice@braindump.be

Hooman Bidgoli
Nokia
Email: hooman.bidgoli@nokia.com

Yisong Liu
China Mobile
Email: liuyisong@chinamobile.com