

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: 4 September 2025

B. Decraene
Orange
K. Talaulikar
K. Raza
Cisco Systems
S. Hegde
Juniper Networks
3 March 2025

SR-MPLS Aggregation Segment
draft-decraene-spring-sr-mpls-aggregation-segment-00

Abstract

One of the key features of IP that has helped IP Routing to scale is aggregation of IP prefixes. This is made possible with longest-match lookup in IP forwarding. Contrary to this, MPLS forwarding works on exact match on MPLS labels. This poses a challenge in aggregation of IP prefixes when the forwarding is based on the MPLS labels associated with those IP prefixes.

This document introduces an Aggregation Segment for Segment Routing with MPLS data plane which can be used to aggregate IP prefixes along with their SR Prefix Segments. Aggregation Segments enable aggregation of IP prefixes to be performed at border routers to improve scalability of MPLS networks.

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 4 September 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. Requirements Language	3
3. Aggregation Segment	3
4. Protocol extensions	3
5. Uses cases	4
5.1. Inter-Areas	4
5.2. Inter-Domain	6
6. Network Operation Considerations	7
7. IANA Considerations	7
8. Security Considerations	7
9. References	7
9.1. Normative References	8
9.2. Informative References	8
Authors' Addresses	9

1. Introduction

One of the key features of IP that has helped IP Routing to scale is aggregation of IP prefixes. This is made possible with longest-match lookup in IP forwarding. IP allows for aggregation between IGP area/levels and/or Autonomous Systems (AS). Optionally, [I-D.ietf-lsr-igp-ureach-prefix-announce] allows for signaling the loss of reachability of an individual prefix covered by the aggregate prefix in order to enable fast convergence mechanisms, e.g., BGP PIC Edge [I-D.ietf-rtgwg-bgp-pic].

Contrary to this, MPLS forwarding works on exact match on MPLS labels. MPLS relies on the propagation of each FEC (viz. a specific IP prefix) across routing areas/domains, either with IGP propagation/redistribution or BGP Label Unicast (BGP-LU) [RFC8277]. This poses a challenge in aggregation of IP prefixes when the forwarding is based on the MPLS labels associated with those IP prefixes.

Moreover, IGP redistribution scales poorly from an IGP and FIB standpoint, while BGP-LU is an additional protocol and infrastructure to deploy and maintain. BGP-LU may also bring additional complexity in some deployments (e.g., PIM rpf-vector, mLDP recursive FEC, BGP local-as when multiple ASes are involved...)

This document introduces an Aggregation Segment for Segment Routing with MPLS (SR-MPLS) data plane which can be used to aggregate IP prefixes along with their SR Prefix Segments. Aggregation Segments provide summarization in the control plane at border routers, and introduce hierarchical MPLS labels in the data plane. The latter allows reduction of scale of FIB entries in the routers of the domains where the aggregate prefix is being used for forwarding.

2. Requirements Language

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.

3. Aggregation Segment

An Aggregation Segment is a segment representing both an aggregate IP prefix and the specific Prefix-SIDs advertised in a compressed way. It forms a two-levels MPLS label hierarchy with the Aggregation Segment at the top level and the specific Prefix Segments at the bottom level.

- * The Aggregation Segment is a Prefix Segment associated with the aggregate IP prefix, and hence it is signaled as a Prefix Segment.
- * The specific Prefix Segments that are covered by the aggregate are signalled in a compressed form, with the advertisement of the first prefix, the absolute label associated with that first prefix, and the number of consecutive prefixes.

The specific prefixes covered by the aggregate prefix MUST all have the same prefix length (e.g., a /32 in case of IPv4 or /128 in case of IPv6).

4. Protocol extensions

The protocol extensions are out of scope of this document but the semantic of the advertisements are defined in this document.

With regards to the signaling of the Aggregation Segment, it is proposed to simply advertise a SR Prefix SID for the IP aggregate prefix.

The advertisement of labels associated with the specific prefixes under the Aggregation Segment requires a protocol extension to advertise the set of prefixes and their label range. This could be signaled in different ways e.g., as sub-TLV of the aggregate prefix advertisement, or as a separate advertisement on the lines of a Segment Routing Mapping Server (SRMS).

If the same Aggregation segment is advertised by multiple nodes, it becomes an anycast segment. Absent specific provisions (e.g., context specific label) such anycast segment needs to advertise the same labels related parameters (first prefix, the absolute label associated with that first prefix) for all instances.

IP routing rules are not modified: routing use the most specific advertisement. The purpose of the Aggregation Segment extension is simply to signal the SID of the Specific Prefix Segment, it does not affect routing.

5. Uses cases

The Aggregation Segment may be used to aggregate multiple Prefix Segments at a routing border such as an ABR for inter-area/levels IGP topologies, or an ASBR for inter-domain topologies.

5.1. Inter-Areas

The following inter-areas diagram is used in this section.

Topology:

```
IS-IS L1 Area 1 | IS-IS L2 | IS-IS L1 Area 2
```

```
PE1 ----- P1 ----- ABR1 ---- P0 ---- ABR2 ----- P2 ----- PE2.3
```

Signaling:

```
=====>
```

```
Aggregation Segment (192.0.2/24)
```

```
- Prefix SID label 2000 (i.e., index 1000)
```

```
----->
```

```
Specific Segment to PE2.x (192.0.2.x)
```

```
- First Prefix 192.0.2.1
```

```
- First Label 1201
```

```
- Range 255
```

LSPs:

```
===== 2000/1203 =====>----- 1203 ----->
```

```
LSP to PE2.3
```

```
===== 2000/1205 =====>----- 1205 ----->
```

```
LSP to PE2.5
```

The addressing schema used in area 2 for egress PE is the following:
PE2.x, IP 192.0.2.x, SID 20x. E.g., for PE2.3, IP 192.0.2.3 and SID 1003. The SR Global Block (SRGB) is 1000.

For scaling purpose, ABR2 advertises a single Aggregation Segment in IS-IS level 2 for all the Prefix Segments of AS2 PEs (PE2.x). ABR1 redistributes this Aggregate Segment into area 1. This Aggregation Segment has a SID 1000 allocated from the SRGB. It also signals a label block for the specific Prefix Segments starting at label 1201 with a range of 255 consecutive labels. Note that, in order to create no additional FIB entry on the aggregation node (ABR) the base label 1201 may be chosen such that the MPLS label used for the specific Prefix SID is the same for the IS-IS L1 Area 2 and for the rest to the network using the Aggregation SID. E.g. for PE2.3, the label used in the area 1 is SRGB + Specific SID = 1000 + 203 and for the rest of the network using the Aggregation Segment, the label used is First Label + Nth Specific SID = 1200 + 3. As a results, all routers in the IGP -except area 2- only install the Global Aggregation Segment SID 1000 with the MPLS label 2000 in their FIB. On an as-needed basis, an ingress PE in the IGP (except in area 2) (e.g., PE1) may install a FIB entry for PE2.3, pushing a two-labels MPLS stack 1003, 2000 (respectively inner and outer labels).

As a result, any ingress PE may reach any egress PE while transit routers only needs to install the single label/forwarding entry of the Aggregation Segment.

5.2. Inter-Domain

The following multi-domain diagram is used in this section.

Topology:

```

      IGP Domain 1      |      IGP Domain 2
      PE1 ----- P1 ----- ASBR ----- P2 ----- PE2.3

```

Signaling:

```

=====>
Aggregation Segment (192.0.2/24)
- SID 1000

```

```

----->
Aggregated Segment to PE2.x (192.0.2.x)
- First Label 1200
- Range 255

```

LSPs:

```

===== 2000/1203 =====>----- 1203 ----->
LSP to PE2.3

```

```

===== 2000/1205 =====>----- 1205 ----->
LSP to PE2.5

```

The addressing schema used in IGP Domain 1 for egress PE is the following: PE2.x, IP 192.0.2.x, SID 20x. E.g., for PE2.3, IP 192.0.2.3 and SID 1003. The SR Global Block (SRGB) is 1000 in IGP Domains 1 and 2. Note that they are independent and do not need to be equal or disjoint.

For scaling purpose, ASBR advertises a single Aggregation Segment in domain 1 for all the Prefix Segments of domain 2 PEs (PE2.x). This global Aggregation Segment has a SID 1000. It also signals a label block for the Specific Prefix Segments starting at label 1200 with a range of 255 consecutive labels. As a results, all routers in domain 1 install a single the Global Aggregation Segment SID 1000 with the MPLS label 2000 in their FIB. On an as-needed basis, ingress PE (PE1) may install a FIB entry for PE2.3, pushing a two-labels MPLS stack 1203, 2000 (respectively inner and outer labels).

As a result, an ingress PE1 in domain 1 may reach any egress PE in domain 2 while transit routers in domain 1 only needs to install the single label/forwarding entry of the Aggregation Segment.

6. Network Operation Considerations

The use of Aggregation Segment reduces the number of states and churn in the control plane protocols, forwarding states and likely configuration, compared to the use of all covered specific Prefix Segments.

Optionally, [I-D.ietf-lsr-igp-ureach-prefix-announce] allows for signaling the loss of reachability of an individual prefix covered by Aggregation Segment in order to enable fast convergence mechanisms, e.g., BGP PIC Edge [I-D.ietf-rtgwg-bgp-pic].

IP Aggregation on border routers requires IP addresses to be aggregatable hence contiguous. Segment aggregation additionally requires that segment ID are allocated in an ordered and contiguous way. This is aligned with the definition and encoding of the Mapping Server as defined in [RFC8667] section 2.4.

In case of redundant ABRs or ASBRs advertising the same Aggregation Segment, it's strongly recommended to use the same SRGB and same Aggregation Segment SID on redundant aggregation nodes, since they are essentially advertising an anycast Prefix-SID. This is not a new requirement compared to the current practice of redistributing each specific Prefix Segments in the IGP.

All ingress which needs to use the Aggregation Segment needs to be upgraded before replacing the redistribution of each Prefix Segment with the Aggregation Segment. A node advertising the Aggregation Segment MAY support the advertisement of both the Aggregation Segment and all specific Prefix-SID in order to ease incremental deployment.

7. IANA Considerations

This document has no IANA actions.

8. Security Considerations

This documents defines a new type of SR-MPLS segments called Aggregation Segment. It does not brings new security considerations. Existing security considerations for Segment Routing and SR-MPLS are described, respectively, in [RFC8402] and [RFC8660].

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/rfc/rfc2119>>.
- [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/rfc/rfc8402>>.
- [RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/rfc/rfc8660>>.
- [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/rfc/rfc8174>>.

9.2. Informative References

- [I-D.ietf-lsr-igp-ureach-prefix-announce] Psenak, P., Filsfils, C., Voyer, D., Hegde, S., and G. S. Mishra, "IGP Unreachable Prefix Announcement", Work in Progress, Internet-Draft, draft-ietf-lsr-igp-ureach-prefix-announce-04, 21 February 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-lsr-igp-ureach-prefix-announce-04>>.
- [I-D.ietf-rtgwg-bgp-pic] Bashandy, A., Filsfils, C., and P. Mohapatra, "BGP Prefix Independent Convergence", Work in Progress, Internet-Draft, draft-ietf-rtgwg-bgp-pic-21, 7 July 2024, <<https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-bgp-pic-21>>.
- [RFC8277] Rosen, E., "Using BGP to Bind MPLS Labels to Address Prefixes", RFC 8277, DOI 10.17487/RFC8277, October 2017, <<https://www.rfc-editor.org/rfc/rfc8277>>.
- [RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <<https://www.rfc-editor.org/rfc/rfc8667>>.

Authors' Addresses

Bruno Decraene
Orange
Email: bruno.decraene@orange.com

Ketan Talaulikar
Cisco Systems
Email: ketant.ietf@gmail.com

Kamran Raza
Cisco Systems
Email: skraza@cisco.com

Shraddha Hegde
Juniper Networks
Email: shraddha@juniper.net