IDR Working Group
Internet-Draft
Intended status: Informational
Expires: 11 March 2025

C. Xie
C. Li
China Telecom
J. Dong
Z. Li
Huawei Technologies
7 September 2024

# Applicability of BGP-LS with Multi-Topology (MT) for Segment Routing based Network Resource Partitions (NRP) draft-ietf-idr-bgpls-sr-vtn-mt-06

## Abstract

Enhanced VPNs aim to deliver VPN services with enhanced characteristics to customers who have specific requirements on their connectivity, such as guaranteed resources, latency, or jitter. Enhanced VPNs require integration between the overlay VPN connectivity and the characteristics provided by the underlay network. A Network Resource Partition (NRP) is a subset of the network resources and associated policies on each of a connected set of links in the underlay network. An NRP could be used as the underlay to support one or a group of enhanced VPN services.

When Segment Routing is used as the data plane of NRPs, each NRP can be allocated with a group of Segment Identifiers (SIDs) to identify the topology and resource attributes of network segments in the NRP. The association between the network topology, the network resource attributes and the SR SIDs may need to be distributed to a centralized network controller. In some network scenarios, each NRP can be associated with a unique logical network topology. This document describes a mechanism to distribute the information of SR based NRPs using BGP-Link State (BGP-LS) with Multi-Topology (MT).

Table of Contents



## 1. Introduction

Enhanced VPNs aim to deliver VPN services with enhanced characteristics to customers who have specific requirements on their connectivity, such as guaranteed resources, latency, or jitter. Enhanced VPNs require integration between the overlay VPN connectivity and the characteristics provided by the underlay network. [RFC9543] discusses the general framework, components, and interfaces for requesting and operating network slices using IETF technologies. A network slice is considered as one target use case of enhanced VPNs.

[RFC9543] also introduces the concept of the Network Resource Partition (NRP), which is a subset of the buffer/queuing/scheduling resources and associated policies on each of a connected set of links in an underlay network. An NRP can be associated with a logical network topology to select or specify the set of links and nodes involved. [I-D.ietf-teas-enhanced-vpn] specifies the framework of NRP-based enhanced VPNs and describes the candidate component technologies in different network planes and network layers. An NRP could be used as the underlay to meet the requirement of one or a group of enhanced VPN services. To meet the requirement of enhanced VPN services, a number of NRPs can be created, each with a subset of network resources allocated on network nodes and links in a customized topology of the physical network.

[I-D.ietf-spring-resource-aware-segments] introduces resource awareness to Segment Routing (SR) [RFC8402]. The resource-aware Segment Identifiers (SIDs) have additional semantics to identify the set of network resources available for the packet processing action associated with the SIDs. As described in [I-D.ietf-spring-sr-for-enhanced-vpn], the resource- aware SIDs can be used to build SR-based NRPs with the required network topology and network resource attributes to support enhanced VPN services. With SR-based data plane, SIDs can be used to represent both the topological instructions and a subset of network resources on the network nodes and links which are allocated to an NRP.

To allow NRP-specific constraint-based path computation and/or NRP-specific shortest path computation to be performed by network controller and network nodes, the set of resource-aware SR SIDs and the associated topology and resource attributes of an NRP need to be distributed using a control plane. When a centralized network controller is used for NRP-specific constraint-based path computation, especially when an NRP spans multiple IGP areas or multiple Autonomous Systems (ASes), BGP-Link State (BGP-LS) [RFC9552] is needed to advertise the NRP information in each IGP area or AS to the network controller, so that the controller could use the collected information to build the view of inter-area or inter-AS SR NRPs.

In some network scenarios, the required number of NRPs could be small, and it can be assumed that each NRP is associated with an independent topology and has a set of dedicated or shared network resources. [I-D.ietf-lsr-isis-sr-vtn-mt] describes the IGP Multi-Topology (MT) [RFC5120] based mechanism to advertise an independent topology and the associated SR SIDs, together with the resource and Traffic Engineering (TE) attributes for each SR based NRP. This document describes a mechanism to distribute the information of SR based NRPs to the network controller using BGP-LS with Multi-Topology.

## 2. Advertisement of Topology Attribute for SR-based NRP

[I-D.ietf-lsr-isis-sr-vtn-mt] describes the IS-IS Multi-Topology based mechanisms to distribute the topology and the SR SIDs associated with SR based NRPs. This section describes the corresponding BGP-LS mechanism to distribute both the intra-domain and inter-domain topology attributes of SR based NRPs.

### 2.1. Intra-domain Topology Advertisement

Section 4.2.2.1 of [RFC9552] defines Multi-Topology Identifier (MT-ID) TLV, which can contain one or more IS-IS or OSPF Multi-Topology IDs. According to [RFC9552], the MT-ID TLV may be present in a Link Descriptor, a Prefix Descriptor, or the BGP-LS Attribute of a Node Network Layer Reachability Information (NLRI).

[RFC9085] defines the BGP-LS extensions to carry the SR-MPLS information using TLVs of BGP-LS Attribute. When Multi-Topology is used with a SR-MPLS data plane, topology-specific Prefix-SIDs and topology-specific Adjacency Segment Identifiers(Adj-SIDs) can be carried in the BGP-LS Attribute associated with the Prefix NLRI and Link NLRI respectively, the MT-ID TLV is carried in the prefix descriptor or link descriptor to identify the corresponding topology of the SIDs.

[RFC9514] defines the BGP-LS extensions to advertise Segment Routing over IPv6 (SRv6) information along with their functions and attributes. When Multi-Topology is used with a SRv6 data plane, the SRv6 Locator TLV is carried in the BGP-LS Attribute associated with the Prefix NLRI, the MT-ID TLV can be carried in the prefix descriptor to identify the corresponding topology of the SRv6 Locator. The SRv6 End.X SIDs are carried in the BGP-LS Attribute associated with the Link NLRI, the MT-ID TLV can be carried in the link descriptor to identify the corresponding topology of the End.X SIDs. The SRv6 SID NLRI is defined to advertise other types of SRv6 SIDs, in which the SRv6 SID descriptors can include the MT-ID TLV so as to advertise topology-specific SRv6 SIDs.

[RFC9552] defines the rules of the usage of MT-ID TLV:

"The MT-ID TLV MAY be included as a Link Descriptor, as a Prefix Descriptor, or in the BGP-LS Attribute of a Node NLRI. When included as a Link or Prefix Descriptor, only a single MT-ID TLV containing the MT-ID of the topology where the link or the prefix is reachable is allowed. In case one wants to advertise multiple topologies for a given Link or Prefix Descriptor, multiple NLRIs MUST be generated where each NLRI contains a single unique MT-ID."

## 2.2. Inter-Domain Topology Advertisement

[RFC9086] and [RFC9514] define the BGP-LS extensions for advertisement of BGP inter-domain topology information and the BGP Egress Peering Segment Identifiers. Such information could be used by a network controller for the computation and instantiation of inter-AS SR TE paths.

In some network scenarios, for instance, an operator's network consists of multiple parts such as metro area networks, backbone networks, or data center networks, each part being a different AS, there are needs to create NRPs which span multiple ASes. The inter-domain NRPs could have different inter-domain connectivity, and may be associated with different sets of network resources in each domain and also on the inter-domain links. In order to build the multi-domain SR based NRPs, it is necessary to advertise the topology and the associated BGP Peering SIDs of each NRP for inter-domain links.

When MT-ID is used consistently in multiple domains covered by an NRP, the topology-specific BGP peering SIDs can be advertised with the MT-ID carried in the corresponding Link NLRI. This can be achieved with the existing mechanisms as defined in [RFC9552][RFC9086] and [RFC9514].

Depending on the requirement of inter-domain NRPs, different mechanisms can be used on the inter-domain connection:

* One External BGP (EBGP) session between two ASes can be established over multiple underlying links. In this case, different underlying links can be used for different inter-domain NRPs, which requires the links to be isolated from each other. In another similar case, the EBGP session is established over a single link, while the network resource (e.g. bandwidth) on this link can be partitioned into several pieces, each of which can be considered as a virtual member link. An NRP can be associated with one of the underlying physical or virtual member links. In both cases, different BGP Peer-Adj-SIDs or SRv6 End.X SIDs need to be allocated to each underlying physical or virtual member link, and the association between the BGP Peer Adj-SID/End.X SID and the MT-ID of the NRP needs to be advertised by the ASBR.

* For inter-domain connection between two ASes, multiple EBGP sessions can be established between different sets of peering ASBRs. It is possible that some of these BGP sessions are used for one inter-domain NRP, while some other BGP sessions are used for another inter-domain NRP. In this case, different BGP Peer Node SIDs need to be allocated to each BGP session and are advertised using the mechanism in [RFC9086] and [RFC9514], the association between the BGP Peer Node SIDs and the MT-ID of the NRP needs to be advertised by the ASBR.

* At the AS-level topology, different inter-domain NRPs may have different inter-AS connectivity. In this case, different BGP Peer Set SIDs are allocated to represent the groups of BGP peers which can be used for load-balancing in each inter-domain NRP. The association between the BGP Peer Node SIDs and the MT-ID of the NRP needs to be advertised by the ASBR.

In network scenarios where consistent usage of MT-ID among multiple domains can not be achieved, a globally-significant identifier many be introduced to identify the inter-domain topology of an NRP. Within each domain, the MT based mechanism could be reused for intra-domain topology advertisement. The detailed mechanism is out of the scope of this document.

## 3. Advertisement of Resource Attribute for SR-based NRP

[I-D.ietf-lsr-isis-sr-vtn-mt] specifies the mechanism to advertise the resource and TE attributes associated with each NRP. This section describes the corresponding BGP-LS mechanisms for reporting NRP resource and TE attributes to network controllers.

The information of the network resources and TE attributes associated with a link of an NRP can be specified by carrying the TE Link attribute TLVs in BGP-LS Attribute [RFC9552], with the associated MT-ID carried in the corresponding Link NLRI.

When the Maximum Link Bandwidth sub-TLV is carried in the BGP-LS attribute associated with the Link NLRI of an NRP, it indicates the amount of link bandwidth resource allocated to the corresponding NRP on the link. The bandwidth allocated to an NRP can be exclusive for traffic in the corresponding NRP. The advertisement of other TE attributes in BGP-LS for NRP is for further study.

## 4. Scalability Considerations

The mechanism described in this document assumes that each NRP is associated with an independent topology, and for the inter-domain NRPs, the MT-IDs used in the involved domains are consistent, so that the MT-IDs can be reused to identify the NRPs in the control plane. Reusing MT-ID can avoid introducing new mechanism with similar functionality in the control plane, while it also has some limitations. For example, even if multiple NRPs share the same topology, each NRP still need to be identified using a unique MT-ID in the control plane. Thus independent path computation needs be executed for each NRP. The number of NRPs supported in a network may be dependent on the number of topologies supported, which is related to both the number of topologies supported in the protocol and the control plane overhead which the network could afford. The mechanism described in this document is considered useful for network scenarios in which the required number of NRPs is small because no control protocol extension is required. For network scenarios where the number of required NRPs is large, more scalable solution would be needed which may require further protocol extensions and enhancements. A detailed analysis about the NRP scalability and the possible optimizations for supporting a large number of NRPs is described in [I-D.ietf-teas-nrp-scalability].

## 5. Security Considerations

This document introduces no additional security vulnerabilities to BGP-LS.

The mechanism proposed in this document is subject to the same vulnerabilities as any other protocol that relies on BGP-LS.

## 6. IANA Considerations

This document does not request any IANA actions.

## 7. Acknowledgments

The authors would like to thank Shunwan Zhuang, Adrian Farrel and Susan Hares for the review and discussion of this document.

## 8. References

### 8.1. Normative References

[I-D.ietf-spring-resource-aware-segments]
Dong, J., Miyasaka, T., Zhu, Y., Qin, F., and Z. Li, "Introducing Resource Awareness to SR Segments", Work in Progress, Internet-Draft, draft-ietf-spring-resource-aware-segments-09, 6 May 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-resource-aware-segments-09>.

[I-D.ietf-spring-sr-for-enhanced-vpn]
Dong, J., Miyasaka, T., Zhu, Y., Qin, F., and Z. Li, "Segment Routing based Network Resource Partition (NRP) for Enhanced VPN", Work in Progress, Internet-Draft, draft-ietf-spring-sr-for-enhanced-vpn-07, 3 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-sr-for-enhanced-vpn-07>.

[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>.

[RFC9085] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Gredler, H., and M. Chen, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing", RFC 9085, DOI 10.17487/RFC9085, August 2021, <https://www.rfc-editor.org/info/rfc9085>.

[RFC9086] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Patel, K., Ray, S., and J. Dong, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering", RFC 9086, DOI 10.17487/RFC9086, August 2021, <https://www.rfc-editor.org/info/rfc9086>.

[RFC9514] Dawra, G., Filsfils, C., Talaulikar, K., Ed., Chen, M., Bernier, D., and B. Decraene, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing over IPv6 (SRv6)", RFC 9514, DOI 10.17487/RFC9514, December 2023, <https://www.rfc-editor.org/info/rfc9514>.

[RFC9552] Talaulikar, K., Ed., "Distribution of Link-State and Traffic Engineering Information Using BGP", RFC 9552, DOI 10.17487/RFC9552, December 2023, <https://www.rfc-editor.org/info/rfc9552>.

## 8.2. Informative References

[I-D.ietf-lsr-isis-sr-vtn-mt]
Xie, C., Ma, C., Dong, J., and Z. Li, "Applicability of IS-IS Multi-Topology (MT) for Segment Routing based Network Resource Partition (NRP)", Work in Progress, Internet-Draft, draft-ietf-lsr-isis-sr-vtn-mt-08, 18 August 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-lsr-isis-sr-vtn-mt-08>.

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Network Resource Partition (NRP) based Enhanced Virtual Private Networks", Work in Progress, Internet-Draft, draft-ietf-teas-enhanced-vpn-20, 14 June 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-enhanced-vpn-20>.

[I-D.ietf-teas-nrp-scalability]
Dong, J., Li, Z., Gong, L., Yang, G., and G. S. Mishra, "Scalability Considerations for Network Resource Partition", Work in Progress, Internet-Draft, draft-ietf-teas-nrp-scalability-05, 5 July 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-nrp-scalability-05>.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <https://www.rfc-editor.org/info/rfc5120>.

[RFC9543] Farrel, A., Ed., Drake, J., Ed., Rokui, R., Homma, S., Makhijani, K., Contreras, L., and J. Tantsura, "A Framework for Network Slices in Networks Built from IETF Technologies", RFC 9543, DOI 10.17487/RFC9543, March 2024, <https://www.rfc-editor.org/info/rfc9543>.

## Authors' Addresses

Chongfeng Xie
China Telecom
China Telecom Beijing Information Science & Technology, Beiqijia

Beijing
102209
China
Email: xiechf@chinatelecom.cn

Cong Li
China Telecom
China Telecom Beijing Information Science & Technology, Beiqijia
Beijing
102209
China
Email: licong@chinatelecom.cn

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Road
Beijing
100095
China
Email: jie.dong@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Road
Beijing
100095
China
Email: lizhenbin@huawei.com

IDR Working Group
Internet-Draft
Intended status: Standards Track
Expires: 30 December 2024

J. Dong
Z. Hu
Huawei Technologies
R. Pang
China Unicom

28 June 2024

# BGP SR Policy Extensions for Network Resource Partition
draft-ietf-idr-sr-policy-nrp-01

## Abstract

Segment Routing (SR) Policy is a set of candidate paths, each consisting of one or more segment lists and the associated information. The header of a packet steered in an SR Policy is augmented with an ordered list of segments associated with that SR Policy. A Network Resource Partition (NRP) is a subset of network resources allocated in the underlay network which can be used to support one or a group of RFC 9543 network slice services.

In networks where there are multiple NRPs, an SR Policy may be associated with a particular NRP. The association between SR Policy and NRP needs to be specified, so that for service traffic which is steered into the SR Policy, the header of the packets can be augmented with the information associated with the NRP. An SR Policy candidate path can be distributed using BGP SR Policy. This document defines the extensions to BGP SR policy to specify the NRP which the SR Policy candidate path is associated with.

Table of Contents



## 1. Introduction

The concept of Segment Routing (SR) policy is defined in [RFC9256]. An SR Policy is a set of candidate paths, each consisting of one or more segment lists. The head end of an SR Policy may learn multiple candidate paths for an SR Policy. The header of a packet steered in an SR Policy is augmented with an ordered list of segments associated with that SR Policy. The BGP extensions to distribute SR Policy candidate paths is defined in [I-D.ietf-idr-sr-policy-safi].

[RFC9543] introduces the concept and the characteristics of RFC 9543 network slice, and describes a general framework for RFC 9543 network slice management and operation. It also introduces the concept Network Resource Partition (NRP), which is a subset of the resources and associated policies in the underlay network. RFC 9543 network slice can be realized by mapping one or more connectivity constructs to an NRP. [I-D.ietf-teas-enhanced-vpn] describes the framework and the candidate component technologies for providing enhanced VPN services based on VPN and Traffic Engineering (TE) technologies.

Enhanced VPN (VPN+) can be used for the realization of RFC 9543 network slices. In the context of network slicing, an NRP is considered as an instantiation of the VTN as defined in [I-D.ietf-teas-enhanced-vpn].

As described in [I-D.ietf-teas-nrp-scalability], one scalable data plane approach to support network slicing is to carry a dedicated NRP ID in the data packet to identify the NRP the packet belongs to, so that the packet can be processed and forwarded using the subset of network resources allocated to the NRP.

In networks where there are multiple NRPs, an SR Policy may be associated with a particular NRP. The association between SR Policy and NRP needs to be specified, so that for service traffic which is steered into the SR Policy, the header of the packets can be augmented with the information associated with the NRP. The association between SR Policy and NRP is described in [I-D.dong-spring-sr-policy-with-nrp]. An SR Policy candidate path can be distributed using BGP SR Policy. This document defines the extensions to BGP SR policy to specify the NRP which the SR Policy candidate path is associated with.

### 1.1. 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 BCP14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

## 2. NRP Identifier of SR Policy

In order to specify the NRP the candidate path of SR policy is associated with, a new sub-TLV called "NRP sub-TLV" is defined in the BGP Tunnel Encapsulation Attribute [RFC9012]. The NRP sub-TLV can be carried in the BGP Tunnel Encapsulation Attribute with the tunnel type set to SR Policy.

The NRP sub-TLV is optional and MUST NOT appear more than once for one SR Policy candidate path. If the NRP sub-TLV appears more than once, the associated BGP SR Policy NLRI is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

The NRP sub-TLV has the following format:

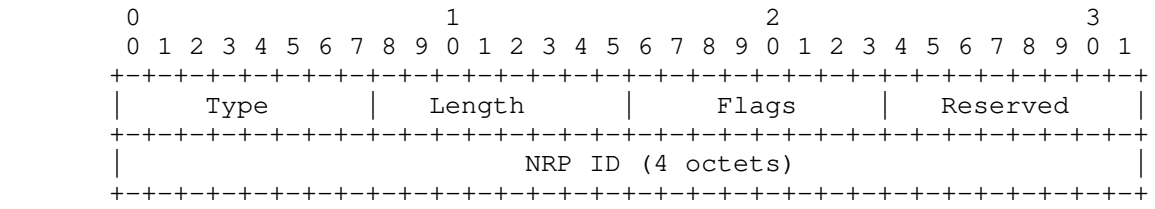



Figure 1. NRP Sub-TLV

where:

* Type: 123

* Length: 6

* Flags: 1-octet flag field. None is defined at this stage. The flags SHOULD be set to zero on transmission and MUST be ignored on receipt.

* RESERVED: 1 octet of reserved bits. It SHOULD be set to zero on transmission and MUST be ignored on receipt.

* NRP ID: A 32-bit domain significant identifier which is used to identify an NRP. Value 0 and 0xFFFFFFFF are reserved.

The encoding structure of BGP SR Policy with the NRP sub-TLV is expressed as below:

```
SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
Attributes:
   Tunnel Encaps Attribute (23)
      Tunnel Type: SR Policy (15)
          Binding SID
          SRv6 Binding SID
          Preference
          Priority
          Policy Name
          Policy Candidate Path Name
          Explicit NULL Label Policy (ENLP)
          NRP
          Segment List
              Weight
              Segment
              Segment
              ...
          ...
```

Figure 2. SR Policy Encoding with NRP sub-TLV

## 3. Procedures

When a candidate path of SR Policy is instantiated within an NRP, and a network-wide data plane NRP ID is used for identifying the resources of the NRP, the originating node of SR Policy SHOULD include the NRP sub-TLV in the BGP Tunnel Encapsulation Attribute of the BGP SR Policy. The setting of other fields and attributes in BGP SR Policy SHOULD follow the mechanism as defined in [I-D.ietf-idr-sr-policy-safi].

On reception of an SR Policy NLRI, a BGP speaker determines if it is acceptable and usable according to the rules defined in Section 4.2 of [I-D.ietf-idr-sr-policy-safi]. If the SR Policy candidate path selected as the best candidate path is associated with an NRP, the headend node of the SR Policy SHOULD encapsulate the NRP ID and the segment list of the selected candidate path in the header of packets which are steered to the SR Policy. For SR Policy with IPv6 data plane, the approach to encapsulate the NRP ID in IPv6 Hop-by-Hop Options header is defined in [I-D.ietf-6man-enhanced-vpn-vtn-id]. For SR Policy with MPLS data plane, one approach to encapsulate the NRP ID to the packet is defined in [I-D.li-mpls-enhanced-vpn-vtn-id].

Although the proposed mechanism allows that different candidate paths in one SR policy be associated with different NRPs, in normal network scenarios it is considered that the association between an SR Policy and NRP is consistent, in such case all candidate paths of one SR policy SHOULD be associated with the same NRP.

## 4. Scalability Considerations

The mechanism specified in this document adds additional information to the SR Policy candidate paths. In order to steer traffic into different NRPs using SR Policy, the SR Policies used for different NRPs need to be different. As the number of NRP increases, the number of SR Policies would also increase accordingly. When BGP is used for distributing SR Policy candidate paths, the amount of control plane information exchanged between the network controller and the headend nodes would also increase. However, since the SR Policies candidate paths distributed in BGP are only installed by the corresponding headend nodes, the impacts to the BGP control plane are considered acceptable.

## 5. Security Considerations

The security considerations of BGP [RFC4271] and BGP SR policy [I-D.ietf-idr-sr-policy-safi] apply to this document.

## 6. IANA Considerations

IANA has assigned the sub-TLV type as defined in Section 2 from "BGP Tunnel Encapsulation Attribute sub-TLVs" registry.

| Value | Description | Reference |
|---|---|---|
| 123 | NRP | This document |

## 7. Acknowledgments

The authors would like to thank Guoqi Xu, Lei Bao, Haibo Wang and Shunwan Zhuang for their review and discussion of this document.

## 8. References

### 8.1. Normative References

[I-D.dong-spring-sr-policy-with-nrp]
Dong, J., Pang, R., and KaZhang, "Associating Segment Routing (SR) Policy with Network Resource Partition (NRP)", Work in Progress, Internet-Draft, draft-dong-spring-sr-policy-with-nrp-00, 27 June 2024, <https://datatracker.ietf.org/doc/html/draft-dong-spring-sr-policy-with-nrp-00>.

[I-D.ietf-idr-sr-policy-safi]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-safi-04, 30 April 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-safi-04>.

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Network Resource Partition (NRP) based Enhanced Virtual Private Networks", Work in Progress, Internet-Draft, draft-ietf-teas-enhanced-vpn-20, 14 June 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-enhanced-vpn-20>.

[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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[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>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[RFC9543] Farrel, A., Ed., Drake, J., Ed., Rokui, R., Homma, S., Makhijani, K., Contreras, L., and J. Tantsura, "A Framework for Network Slices in Networks Built from IETF Technologies", RFC 9543, DOI 10.17487/RFC9543, March 2024, <https://www.rfc-editor.org/info/rfc9543>.

## 8.2. Informative References

[I-D.ietf-6man-enhanced-vpn-vtn-id] Dong, J., Li, Z., Xie, C., Ma, C., and G. S. Mishra, "Carrying Network Resource Partition (NRP) Information in IPv6 Extension Header", Work in Progress, Internet-Draft, draft-ietf-6man-enhanced-vpn-vtn-id-06, 20 February 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-6man-enhanced-vpn-vtn-id-06>.

[I-D.ietf-teas-nrp-scalability] Dong, J., Li, Z., Gong, L., Yang, G., and G. S. Mishra, "Scalability Considerations for Network Resource Partition", Work in Progress, Internet-Draft, draft-ietf-teas-nrp-scalability-04, 4 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-nrp-scalability-04>.

[I-D.li-mpls-enhanced-vpn-vtn-id]
Li, Z. and J. Dong, "Carrying Virtual Transport Network (VTN) Information in MPLS Packet", Work in Progress, Internet-Draft, draft-li-mpls-enhanced-vpn-vtn-id-03, 16 October 2022, <https://datatracker.ietf.org/doc/html/draft-li-mpls-enhanced-vpn-vtn-id-03>.

## Authors' Addresses

Jie Dong
Huawei Technologies
Email: jie.dong@huawei.com

Zhibo Hu
Huawei Technologies
Email: huzhibo@huawei.com

Ran Pang
China Unicom
Email: pangran@chinaunicom.cn

Network Working Group
Internet Draft
Intended status: Standards Track
Expires: January 9, 2025

C. Lin
M. Chen
H. Li
New H3C Technologies

July 9, 2024

# BGP-LS Advertisement of TE Policy Performance Metric
# draft-lin-idr-bgpls-te-policy-pm-03

## Abstract

This document describes a way to advertise the performance metrics for Traffic Engineering (TE) Policy using BGP Link State (BGP-LS).

## 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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on January 9, 2025.

## Copyright Notice

Table of Contents



## 1. Introduction

BGP Link State (BGP-LS) can be used to distribute link-state and traffic engineering (TE) information to external components [RFC7752]. [I-D.ietf-idr-bgp-ls-te-path] describes the mechanism for BGP-LS to distribute the information of TE policies. [I-D.ietf-idr-bgp-ls-sr-policy] describes the mechanism for BGP-LS to distribute the information of SR policies.

In some network scenarios, the controller needs to obtain the performance information of TE Policies, which can be used in service placement to meet better customer requirements and utilize network resources more efficiently.

This document describes a way to advertise the performance metrics for Traffic Engineering (TE) Policy using BGP Link State (BGP-LS).

### 1.1. 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.

## 2. Advertisement of TE Policy Performance Metric

[RFC8571] defines several Link Attribute TLVs for BGP-LS to carry the IGP Traffic Engineering Performance Metric Extensions:

| TLV Code Point | Value |
|---|---|
| 1114 | Unidirectional Link Delay |
| 1115 | Min/Max Unidirectional Link Delay |
| 1116 | Unidirectional Delay Variation |
| 1117 | Unidirectional Link Loss |
| 1118 | Unidirectional Residual Bandwidth |
| 1119 | Unidirectional Available Bandwidth |
| 1120 | Unidirectional Utilized Bandwidth |

The above TLVs can be re-used to advertise the performance metrics for TE Policies.

When used to describe the performance metric of the TE Policy NLRI, they are carried in the optional non-transitive BGP Path Attribute "BGP-LS Attribute" defined in [RFC7752]. The semantics of the above TLVs comply with [RFC8571], except for that they are extended to describe TE Policies besides IGP links.

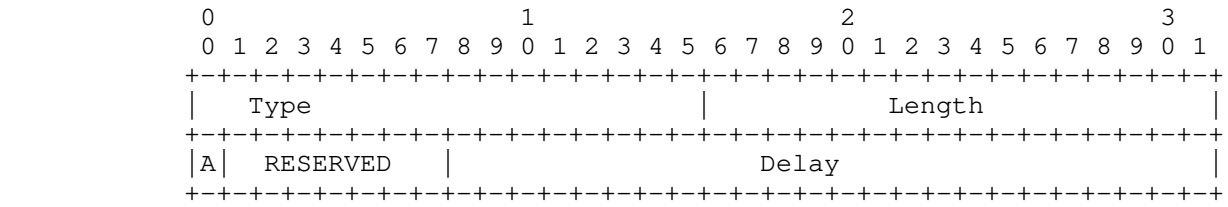
The performance metric of TE Policy may be measured at the headend, for example, by using TWAMP for SR Policy. But the measurement methods are out of the scope of this document.

The existing performance metrics above are all unidirectional. However, there are also requirements to advertise round-trip performance metrics for TE Policies. The BGP-LS extensions for round-trip TE performance metrics are defined in the following section.

## 3. Extensions for Round-trip TE Performance Metric

### 3.1. Round-trip Delay TLV

This TLV advertises the average round-trip delay for TE Policy.

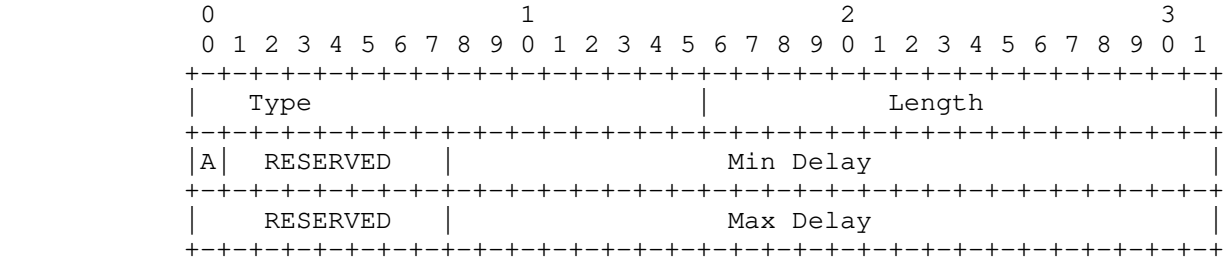



where:

- Type: TBD
- Length: 4
- Reserved: Reserved for future use. MUST be set to 0 when sent and MUST be ignored when received.
- A: Anomalous (A) Bit. Same with the A Bit in Unidirectional Link Delay TLV [RFC8571].
- Delay: Similar with the Delay filed in Unidirectional Link Delay TLV [RFC8571], except for that the delay is round-trip.

### 3.2. Min/Max Round-trip Delay TLV

This TLV advertises the minimum and maximum round-trip delay for TE Policy.




where:

- Type: TBD
- Length: 4
- Reserved: Reserved for future use. MUST be set to 0 when sent and MUST be ignored when received.

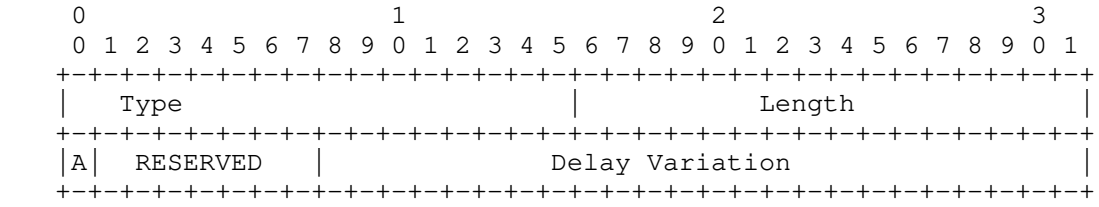
- o A: Anomalous (A) Bit. Same with the A Bit in Min/Max Unidirectional Link Delay TLV [RFC8571].

- o Min Delay: Similar with the Min Delay filed in Min/Max Unidirectional Link Delay TLV [RFC8571], except for that the delay is round-trip.

- o Max Delay: Similar with the Max Delay filed in Min/Max Unidirectional Link Delay TLV [RFC8571], except for that the delay is round-trip.

### 3.3. Round-trip Delay Variation TLV

This TLV advertises the average round-trip delay variation for TE Policy.




where:

- o Type: TBD

- o Length: 4

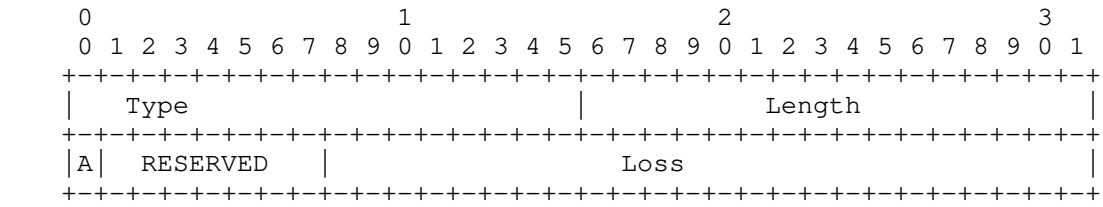
- o Reserved: Reserved for future use. MUST be set to 0 when sent and MUST be ignored when received.

- o A: Anomalous (A) Bit. Same with the A Bit in Unidirectional Delay Variation TLV [RFC8571].

- o Delay Variation: Similar with the Delay Variation filed in Unidirectional Delay Variation TLV [RFC8571], except for that the delay variation is round-trip.

### 3.4. Round-trip Loss TLV

This TLV advertises the round-trip loss for TE Policy.




where:

- Type: TBD
- Length: 4
- Reserved: Reserved for future use. MUST be set to 0 when sent and MUST be ignored when received.
- A: Anomalous (A) Bit. Same with the A Bit in Unidirectional Link Loss TLV [RFC8571].
- Loss: Similar with the Link Loss filed in Unidirectional Link Loss TLV [RFC8571], except for that the loss is round-trip.

## 4. Security Considerations

This document does not introduce additional security issues than those described in [RFC7752], [I-D.ietf-idr-bgp-ls-te-path] and [I-D.ietf-idr-bgp-ls-sr-policy].

## 5. IANA Considerations

This document defines the following TLVs for BGP-LS.

| TLV Code Point | Value |
|---|---|
| TBD | Round-trip Delay |
| TBD | Min/Max Round-trip Delay |
| TBD | Round-trip Variation |
| TBD | Round-trip Loss |

## 6. References

### 6.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, May 2017

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC8571] Ginsberg, L., Ed., Previdi, S., Wu, Q., Tantsura, J., and C. Filsfils, "BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions", RFC 8571, DOI 10.17487/RFC8571, March 2019, <https://www.rfc-editor.org/info/rfc8571>.

[I-D.ietf-idr-bgp-ls-te-path] Previdi, S., Talaulikar, K., Dong, J., Gredler, H., and J. Tantsura, "Advertisement of Traffic Engineering Paths using BGP Link-State", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-te-path-01, September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ls-te-path-01>.

[I-D.ietf-idr-bgp-ls-sr-policy] Previdi, S., Talaulikar, K., Dong, J., Gredler, H., and J. Tantsura, "Advertisement of Segment Routing Policies using BGP Link-State", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-sr-policy-03, November 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ls-sr-policy-03>.

Authors' Addresses

Changwang Lin
New H3C Technologies

Email: linchangwang.04414@h3c.com

Mengxiao Chen
New H3C Technologies

Email: chen.mengxiao@h3c.com

Hao Li
New H3C Technologies

Email: lihao@h3c.com

IDR Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 05, 2025

Y. Liu
China Mobile
C. Lin
New H3C Technologies
Ran.Chen
ZTE
Y. Qiu
New H3C Technologies

September 06, 2024

# BGP Extension for SR Segment List optimization
draft-liu-idr-sr-segment-list-optimize-00

## Status of this Memo



## Copyright Notice

Abstract

This document introduces an optimization method for segment list arrangement the SR Policy TLV of the BGP Tunnel Encapsulation Attribute. This optimization solves the problem of the Segment Routing's penultimate segment node being unable to perform the penultimate Segment Pop (PSP) behavior when the egress node has both End SID and service SID encapsulated in Segment Routing Header's segment List. This solution adds an E-Flag to the SRv6 SID Endpoint Behavior sub-TLV carried in Segment List Sub-TLV of the SR Policy TLV.

This optimization can improve the forwarding efficiency of data packets when End SID and Service SID are present.

Table of Contents

## 1. Introduction

Segment Routing (SR) [RFC8402] allows a headend node to steer a packet flow along any path. Intermediate per-path states are eliminated thanks to source routing.

The headend node is said to steer a flow into an SR Policy [RFC8402]. The packets steered into an SR Policy carry an ordered list of segments associated with that SR Policy.

[I-D. draft-ietf-idr-sr-policy-safi] specifies how BGP may be used to distribute SR Policy candidate paths. New sub-TLVs for the Tunnel Encapsulation Attribute are defined for signaling information about these candidate paths.

This document introduces an optimization method for segment list arrangement to solve the problem of the penultimate segment node being unable to perform PSP behavior when the egress node has both End SID and service SID, and improve the forwarding efficiency of data packets.

## 2. Terminology

The following terminologies are used in this document.

SR: Segment Routing

SRv6: SR for IPv6

SRH: Segment Routing Header

SID: Segment Identifier

CE: Customer Edge

PE: Provider Edge

VPN: Virtual Private Network

PSP: Penultimate Segment Pop

## 3. Requirement background

In SRv6 networks, some functions can only be executed on the penultimate SR Segment Endpoint Node, such as Penultimate Segment Pop (PSP) behavior. However, if both the End SID and service SID of the egress node are encapsulated in SRH.SegmentList, the endpoint

will not be able to identify itself as the penultimate SR Segment Endpoint Node based on the SRH.SL field after receiving the packet.

For example, in the following scenarios, the Segment List of SRv6 Policy must include the End SID of the egress node. The SRH extension header of VPN user's data packets forwarded based on this SRv6 Policy tunnel will simultaneously encapsulate the End SID and VPN SID of the egress node.

* Scenario 1

  In tunnel splicing scenarios and cross domain path splicing scenarios, usually based on binding SID to steer traffic. The Segment List of SRv6 Policy on the head node must include the End SID of the egress node.

* Scenario 2

  When the head node enables end-to-end fast fault detection of SRv6 Policy, OAM messages are sent to the egress node. The End SID of the egress node must be specified in the Segment List of this SRv6 Policy.

In this way, the following two problems will arise:

* Problem 1: PSP behavior may not be executable.

  If the head node encapsulates both the End SID and VPN SID of the egress node in the SRH.SegmentList, the penultimate SR Segment Endpoint Node will find that local SID is not in the position with SL=1 after receiving the packet.

  After executing SL--, SL is still greater than 0. Because the condition of (SL==0) is not met, the penultimate SR Segment Endpoint Node will not be able to perform the processing of removing the SRH from the IPv6 extension header.

* Problem 2: The forwarding efficiency of egress node decreases.

  If the egress node receives a packet with both a local End SID and a VPN SID, it needs to first look up the table based on the End SID. Then, based on the VPN SID, execute the VPN SID instruction, and finally remove the outer IPv6 packet header and forward it to VPN network.

  The data packet needs to look up the SID table twice within the egress node. For some chips, the second SID table lookup requires a loopback interface to be implemented. Due to the bandwidth

limitations and the possibility of other service packets coexisting on the loopback interface, the forwarding efficiency of packets to VPN will be greatly affected.

* Problem 3: Increase the overhead of the packet header.

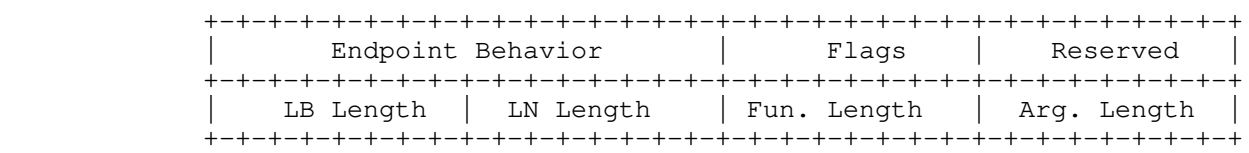
Carrying both the End SID and VPN SID of the egress node in the SRH.SegmentList will increase the overhead of the packet header. Especially in environments that require SRv6 header compression, arranging End SID for egress node will reduce compression efficiency.

Therefore, this document proposes a method to optimize the SRH.SegmentList encapsulated by the head node. When there are End SID and service SID of egress node on the path at the same time, only the service SID is encapsulated in the SRH.SegmentList.

This can solve the problem of the penultimate segment node being unable to perform PSP behavior when the egress node has both End SID and service SID, and improve the forwarding efficiency of data packets on the egress node.

4. Extend the Reserved field of SRv6 SID Endpoint Behavior and Structure

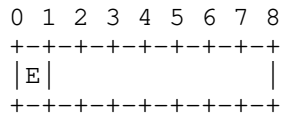
Extend the Reservied field of SRv6 SID Endpoint Behavior and Structure defined in Chapter 2.4.4.2.4 of[I-D.ietf-idr-sr-policy-safi], Define a bit to identify whether this SID belong to the egress node.




where:

Flags: 1 octet of flags. It appears the SRv6 SID Endpoint Behavior and Structure is on segment types B, I, J, and K, which specified in [I-D.draft-ietf-idr-sr-policy-safi] and [I-D.draft-ietf-idr-sr-segtypes-ext].

o




* E-Flag: This flag, when set, indicates that this segment is the egress node's SID.

## 5. Optimizing the arrangement method of segment list

After the controller arranges the SRv6 forwarding path, it informs the ingress node which is the egress node's SID through the E-Flag.

When the controller distributes the SRv6 Policy configuration to the head node through BGP, the E-Flag bit of SRv6 SID Endpoint Behavior and Structure in the segment sub-TLV corresponding to the egress node is set to 1. And the E-Flag bits corresponding to the ingress node and intermediate node are set to 0.

After receiving the SRv6 Policy configuration with E-Flag, the ingress node will not simultaneously arrange the End SID and Service SID of the egress node into the SRH.SegmentList of packet.

For data packets forwarded to VPN through this SRv6 Policy, the SRH.SegmentList will not encapsulate the End SID corresponding to the egress node in the SID list of SRv6 Policy.

If the forwarding path does not include the service SID of the egress node, then the End SID of the egress node should be encapsulated in SRH.SegmentList.

For OAM detection packets of the SR policy, the SRH.SegmentList is encapsulated according to the SID list of the SR policy, only encapsulating node SIDs.

## 6. Example of SRv6 packet Processing Process

Taking Figure 1 as an example, describe how SRv6 data packets and OAM packets are forwarded in the SRv6 network based on the optimized Segment List arrangement mechanism.

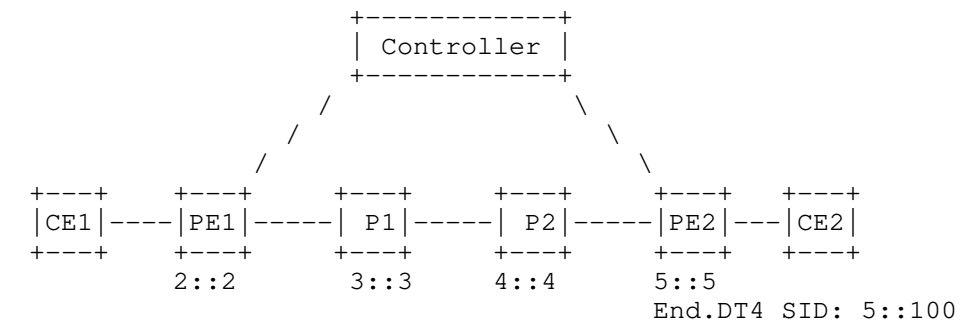



Figure 1

CE1 and CE2 are VPN access devices that connect to the IPv6 backbone network through PE. PE1 has a locator 2::/64. P1 has a locator 3::/64. P2 has an End SID 4::4 with PSP Flavor. PE2 has a locator 5::/64 and a VPN SID 5::100. The traffic from CE1 to CE2 is forwarded along the path PE1->P1->P2->PE2.

P2 needs to perform the PSP behavior to remove the SRH extension header.

The controller calculates the SRv6 forwarding path from PE1 to PE2 based on the collected topology and configuration information, and distributes the SRv6 Policy to PE1 through BGP. The Endpoint address is 5::5 of PE2. There is only one candidate path. The candidate path contains a Segment list <3::3, 4::4, 5::5>. For PE2's Segment 5::5, the E-Flag bit of SRv6 SID Endpoint Behavior and Structure in the segment sub-TLV set to 1.

PE2 advertises a BGP VPN route to PE1, and the next hop of the BGP route is the endpoint address 5::5. After receiving the BGP route, PE1 iterates to the SRv6 Policy using the color and the next hop of the route.

There are two types of packets sent from PE1 to PE2: data packets and OAM packets.

### 6.1. Data packet Processing to VPN

After PE1 receives the data packet from CE1 to CE2, it looks up the VPN instance routing table and iterates to SRv6 Policy.

PE1 adds the SRH extension header to the packet and encapsulates the Segment List of the SRv6 Policy. The Segment List in the SRH extension header is encapsulated as <3::3, 4::4, 5::100>, and the SL is set to 2.

The Segment List in SRH is shown in Figure 2.

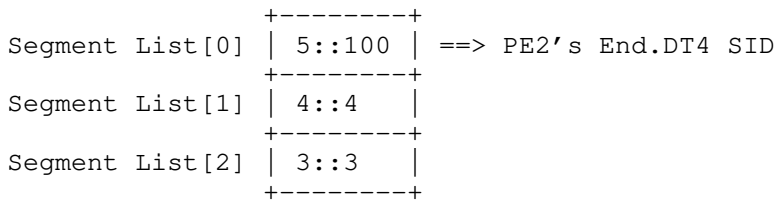



Figure 2

The segment list optimization method proposed in this document is suitable for both SRv6 SID compressed and non-compressed scenarios. If the END SID and VPN SID of the egress node share a common Locator-Block with a sequence of consecutive nodes, the SIDs of the egress node can also be arranged in a compressed Segment List.

In order to improve compression efficiency and reduce the overhead of SRv6 packet header, the compressed Segment List can only contain the compressed VPN SID.

As shown in Figure 3, PE1, P1, P2, and PE3 share the common Locator-block A:0:0:0/64 (represented by LB in Figure3).

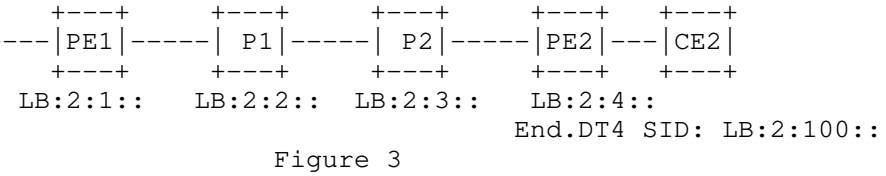



Figure 3

The compressed Segment List optimized in SRH is shown in Figure 4.

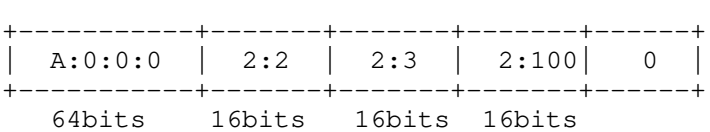



Figure 4

6.2. OAM Packet Processing to the Egress Node

If the head node enables OAM function and detects a fault in the SRv6 Policy forwarding path, PE1 will send OAM detection messages to PE2, such as BFD packets.

The OAM detection message sends by PE1 encapsulate the segment list corresponding to the SRv6 Policy. Since the message does not need to

be sent to VPN, the Segment List of the SRH extension header is encapsulated as <3::3, 4::4, 5::5>.

The Segment List in SRH is shown in Figure 5.

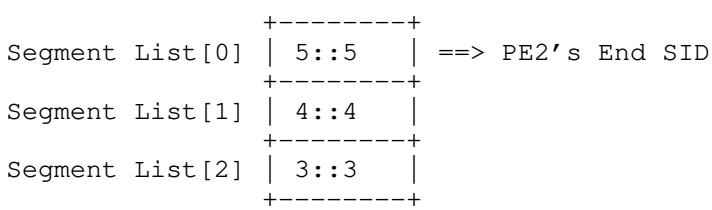



Figure 5

## 7. IANA Considerations

No requirements for IANA.

## 8. Security Considerations

[RFC8754] defines the notion of an SR domain and use of SRH within the SR domain. The use of egress protection mechanism described in this document is restricted to an SR domain. Procedures for securing an SR domain are defined the section 5.1 and section 7 of [RFC8754].

This document does not impose any additional security challenges to be considered beyond security threats described in [RFC8754], [RFC8679] and [RFC8986].

## 9. References

### 9.1. Normative References

[I-D.ietf-idr-sr-policy-safi]Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-safi-06, 26 July 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-safi-06>.

[I-D.ietf-idr-bgp-sr-segtypes-ext] Talaulikar, K., Filsfils, C., Previdi, S., Mattes, P., and D. Jain, "Segment Routing Segment Types Extensions for BGP SR Policy", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-sr-segtypes-ext-04, 30 July 2024,<https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-sr-segtypes-ext-04>.

[RFC8400] Chen, H., Liu, A., Saad, T., Xu, F., and L. Huang, "Extensions to RSVP-TE for Label Switched Path (LSP) Egress Protection", RFC 8400, DOI 10.17487/RFC8400, June 2018, <https://www.rfc-editor.org/info/rfc8400>.

[RFC8679] Shen, Y., Jeganathan, M., Decraene, B., Gredler, H., Michel, C., and H. Chen, "MPLS Egress Protection Framework", RFC 8679, DOI 10.17487/RFC8679, December 2019, <https://www.rfc-editor.org/info/rfc8679>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header(SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[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>.

### 9.2. Informative References

TBD

## 10. Acknowledgments

TBD

## Authors' Addresses

Yisong Liu
China Mobile

Email: liuyisong@chinamobile.com

Changwang Lin
New H3C Technologies

Email: linchangwang.04414@h3c.com

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.cn

Yuanxiang Qiu
New H3C Technologies

Email: qiuyuanxiang@h3c.com