rtgwg Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 17, 2024

W. Cheng
W.Jiang
China Mobile
C. Lin
New H3C Technologies
Z. Hu
Huawei Technologies
Y. Qiu
New H3C Technologies
September 17, 2023

SRv6 Egress Protection in Multi-homed scenario
draft-cheng-rtgwg-srv6-multihome-egress-protection-05

Status of this Memo

Copyright Notice

Abstract

This document describes a SRv6 egress node protection mechanism in multi-homed scenarios.

Table of Contents

1. Introduction

The fast protection of a transit node of a Segment Routing (SR) path or tunnel is described in [I-D.ietf-rtgwg-segment-routing-ti-lfa] and [I-D.hu-spring-segment-routing-proxy-forwarding]. [RFC8400] specifies the fast protection of egress node(s) of an MPLS TE LSP tunnel including P2P TE LSP tunnel and P2MP TE LSP tunnel in details. However, these documents do not discuss the fast protection of the egress node of a Segment Routing for IPv6 (SRv6) path or tunnel.

[I-D.ietf-rtgwg-srv6-egress-protection] proposes mirror protection mechanism and presents protocol extensions for the fast protection of the egress node of a SRv6 path or tunnel. However, the mechanism provided in this document is relatively complex. It is necessary to configure the Mirror SID for the protected egress node on the backup egress node. The mirror relationship needs to be distributed through IGP and BGP protocols to automatically create mapping entries.

This document introduces a simplified protection mechanism of the egress node of a SRv6 path. Only expanding the data plane can perform fast path switching in case of egress node failure.

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

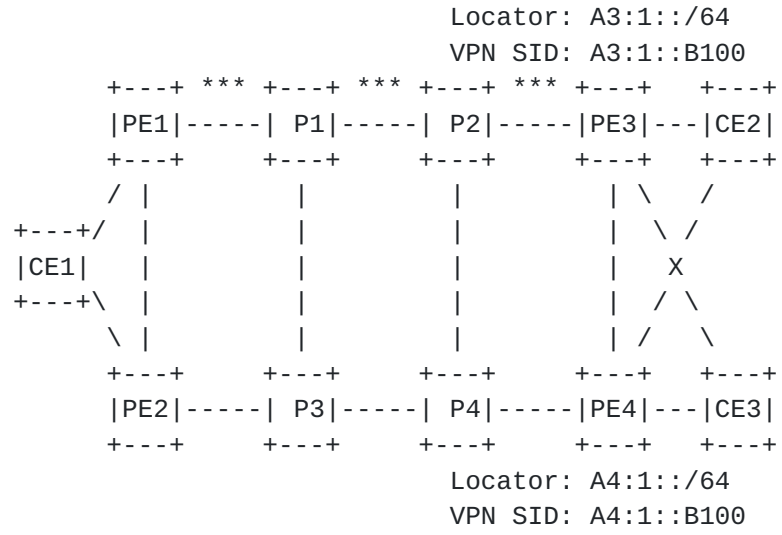
VPN: Virtual Private Network

PSD: Penultimate Segment Decapsulation

3. Multi-homed SRv6 Egress Protection Mechanism

This section describes the mechanism of SRv6 path egress protection in multi-homed scenarios and the extension of SRH extension header.

Figure 1 is used to explain the multi-homed egress node protection mechanism.



PE3 Egress
PE4 Backup Egress
CEx Customer Edge
Px Non-Provider Edge
*** SR Path

Figure 1

3.1. Procedure on the Ingress Endpoint

In the CE multi-homed or double-homed scenario, the ingress node learns the multi-homed or double-homed route of CE through the routing protocol, and then determines the optimal path and suboptimal path according to the routing optimization strategy. The egress node on the optimal path acts as the primary egress, and the SID of the primary egress node is used as the primary SID. The egress node on the suboptimal path acts as the backup egress, and the SID of the backup egress node is used as the backup SID.

On the path forwarded based on SRv6 TE policy, when the ingress node encapsulates the SRH extension header, judge whether the primary VPN SID of the egress node (PE1) has a backup SID. If yes, insert the backup SID into the position of Segment List[0]. The format of SRH extension header filling is shown in the following Figure 2.

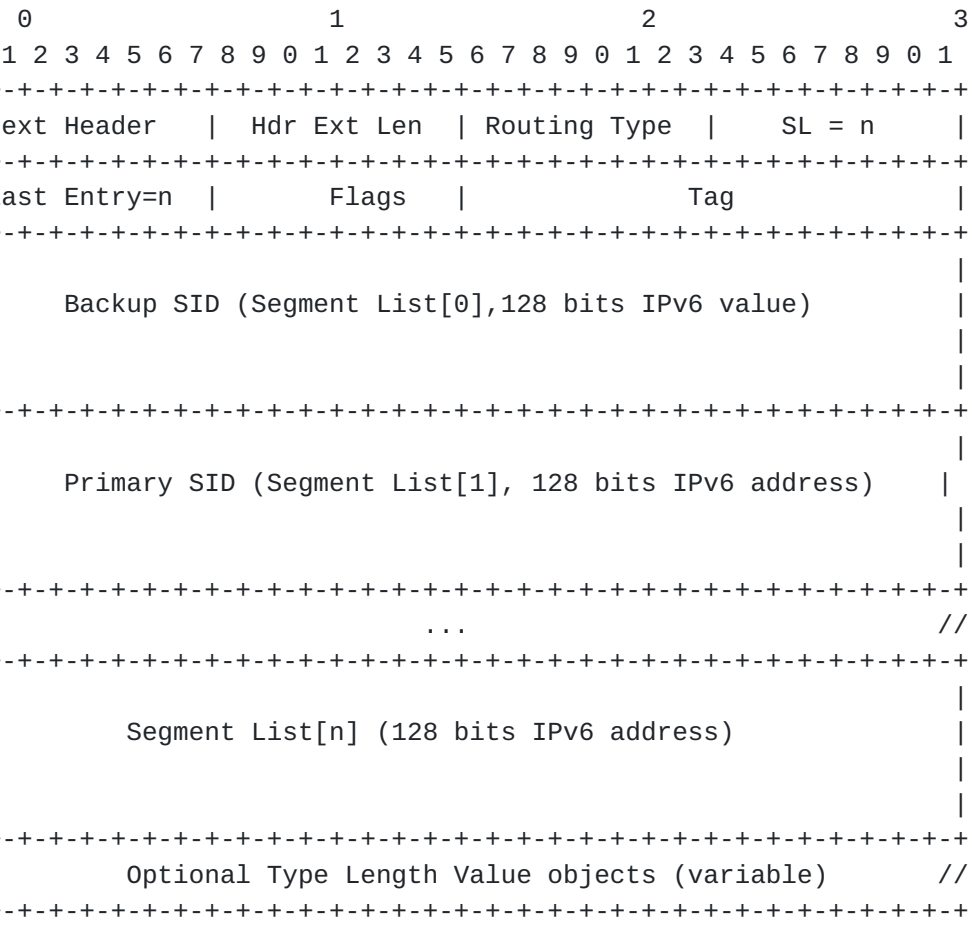


Figure 2

3.2. Procedure on the Egress Endpoint

Normally, the traffic is forwarded along the path P1->P2->PE3->CE2. When the primary egress node PE3 receives a packet whose IPv6 DA is a local SID, PE3 removes the outer packet header with all its extension headers and submits the packet to the egress FIB lookup for transmission to the new destination. However, because Segment

List[0] carries the backup SID, the SID of the primary egress node PE3 is encapsulated in the position of the penultimate SID. Therefore, according to [RFC8986], PE3 will not perform SRv6 decapsulation.

In order to indicate the primary egress node with SID located in the penultimate position of the SRH Segment List array to decapsulate, this document define PSD(Penultimate Segment Decapsulation) Flavor.

SR Segment Endpoint nodes receive the IPv6 packet with the Destination Address field of the IPv6 header equal to its SID address.

When a node (PE3 in Figure 1) receives a packet whose IPv6 DA is a SID with PSD Flavor located in the penultimate position of the SRH Segment List array and that SID is a local SID, it indicates to remove the outer encapsulation of the packet, and forward the packet according to the exposed packet.

PSD Flavor can apply to End.DT4, End.DT6, End.DT46, End.DX4, End.DX6, End.DX2, End.DX2V and End.DX2M. The SIDs can be advertised via routing protocol.

The SRH processing of the End.DT4, End.DT6, End.DT46, End.DX4, End.DX6, End.DX2, End.DX2V and End.DX2M behaviors defined in [RFC8986] are modified; the instructions S02 are substituted by the following one:

```
S02.     If ((Segments Left != 0) && (Segments Left != 1)) {
```

Due to the above pseudocode modification, the PSD operation only takes place at the egress node and does not happen at any transit node. When a SID of PSD flavor is processed at a transit node, the PSD behavior is not performed since Segments Left would not be 1 or 0.

Normally, the traffic is forwarded along the path P1->P2->PE3->CE2. When PE3 receives a packet whose IPv6 DA is S and S is a local PSD-flavored SID, PE3 removes the outer packet header with all its extension headers and submits the packet to the egress FIB lookup for transmission to the new destination.

3.3. Procedure on the Penultimate Endpoint

After receiving the packet, if any of the following cases is met, the penultimate endpoint acting as the repair node can provide fast

protection against the failure of the egress node by IGP fast convergence or BFD detection.

* Case 1: When the destination address of the received packet is the local END.X SID, the outgoing interface is down. Update the IPv6 DA with SRH[SL] and the SL is reduced.

* Case 2: When the destination address of the received packet is the local END SID, the FIB is looked up for the updated DA with Segment List[SL] and the SL is reduced.

IF the primary outbound interface used to forward the packet failed or there is no FIB entry for forwarding the packet, the detailed processing to be performed by the penultimate node is as follows:

```
IF SL = 1 THEN
   SL decreases by 1 and becomes 0;
   Update the IPv6 DA with Segment List[0];
   FIB lookup on the updated DA;
   Forward the packet according to the matched entry;
```

When primary egress node (PE3) fails, the Penultimate Endpoint (P2) finds out that the PE3's SID is unreachable. For example, P2 quickly detects that the route to PE3 is unreachable through IGP convergence or BFD detection. Then, P2 sequentially looks up the Segment List after the current unreachable SID and finds the first reachable downstream nodes. The specific operations are as follows:

P2 decreases SL by 1 and then modifies the destination address of the packet to Segment List[SL]. Because Segment List[0] is the backup SID, if the backup egress node (PE4) is reachable, P2 sends the modified packet to PE4. In this way, P2 provides fast protection for the egress failure and greatly shortens the cut-off time.

When the packet arrives at PE4, PE4 removes the outer IPv6 header, and forwards the original packet.

After the route convergence is completed, the ingress node (PE1) will reselect the forwarding path for the traffic to VPN, and switch the path (P1->P3->P4->PE4->CE2) to the CE to the egress node (PE4). After that, P2 no longer needs to forward the packet with the destination address of PE3's PSD-flavored SID.

About penultimate node, it could be several hops away to egress node, according to [RFC8986], the transit node cannot process SRH. If there are transit node(s) between the penultimate endpoint node and the primary egress node, the multi-homed SRv6 egress protection mechanism does not take effect on the transit node. The multi-homed

egress protection is only performed at the penultimate endpoint node.

4. Multi-homed SRv6 Egress Protection Example

Figure 3 shows an example of protecting egress PE3 of a SRv6 TE path, which is from ingress PE1 to egress PE3.

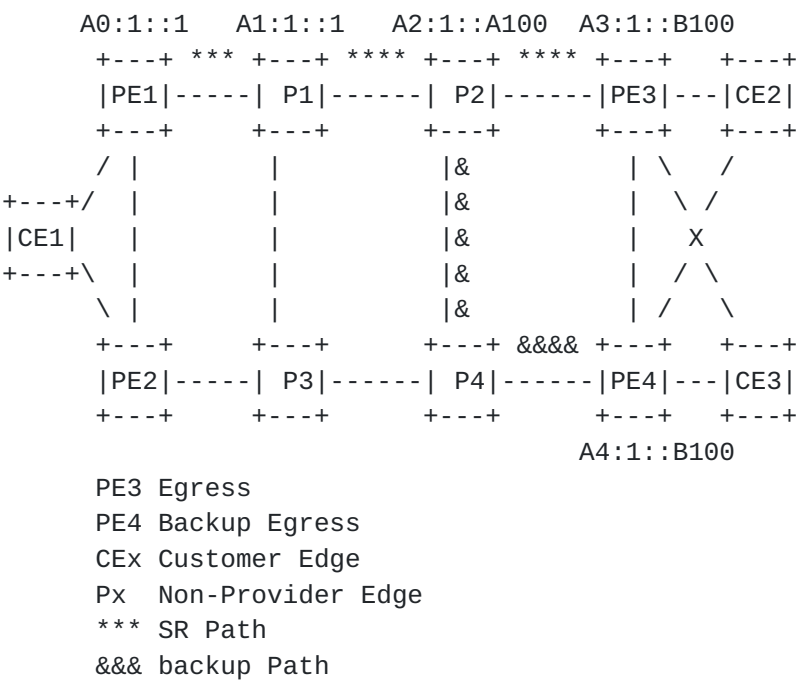


Figure 3

In this document, a SID list is represented as <S1, S2, S3> where S1 is the first SID to visit, S2 is the second SID to visit and S3 is the last SID to visit along the SRv6 path.

In Figure 3, Both CE2 and CE3 are dual homed to PE3 and PE4. PE1 has a locator A0:1::/64. P1 has a locator A1:1::/64. P2 has a locator A2:1::/64 and END.X SID A2:1::A100. PE3 has a locator A3:1::/64 and a VPN SID A3:1::B100 with PSD-flavored behavior. PE4 has a locator A4:1::/64 and VPN SID A4:1::B100. The traffic from CE1 to CE2 is forwarded along the path PE1->P1->P2->PE3.

After the configuration, according to the BGP route selection principle, the ingress PE node selects the preferred route as the primary node and the second-best route as the backup node. PE1 determines that PE3 is the primary egress node and PE4 is the backup egress node. PE3's backup SID is PE4's VPN SID.

After multi-homed egress protection is enabled, when PE1 receives the packet from CE1 to CE2, PE1 encapsulates the packet with IPv6

header. The segment list in SRH is designed as < A1:1::1, A2:1::A100, A3:1::B100, A4:1::B100>. The SL is set to 3.

In normal operations, When P2 receives a packet destined to END.X SID A2:1::A100, it decreases SL by 1 and forwards the packet with source A0:1::1 and destination PE3's VPN SID A3:1::B100 from the link between P2 and PE3 according to END.X SID. The SL of the packet sent to PE3 is 1. When the packet arrives at PE3, the destination address A3:1::B100 matches PE3's PSD-flavored SID. PE3 as the penultimate endpoint performs decapsulation and forwarding processing. The specific operations of PE3 are as follows:

1) Remove the outer packet header and all its extension headers.

2) Look up the FIB table according to the destination address of the original packet.

3) Send the packet to CE2 according to the FIB entry.

After PE3 fails, P2 detects PE3 failure through IGP fast convergence or BFD detection. When P2 receives the packet to be sent to PE3's VPN SID A3:1::B100, after decreasing SL P2 finds that the outgoing interface is down or the route to PE3 is unreachable. P2 continues to decrease SL by 1 and obtains the next reachable SID of PE4's SID from the segment list. The SL changes to 0. P2 changes the destination address of the packet with the backup SID of Segment List[0] and sends the modified packet to A4:1::B100.

When PE4 receives the modified packet, it decapsulates the packet and forwards the decapsulated packet by executing END.DT6 behavior for an END.DT6 SID instance.

5. IANA Considerations

This document requests IANA to allocate the following codepoints for PSD flavor behaviors within the "SRv6 Endpoint Behaviors" registry in the "Segment Routing registry group.

Value	Hex	Endpoint behavior	Reference
140	0x008C	End.DX6 with PSD	[This.ID]
141	0x008D	End.DX4 with PSD	[This.ID]
142	0x008E	End.DT6 with PSD	[This.ID]
143	0x008F	End.DT4 with PSD	[This.ID]
144	0x0090	End.DT46 with PSD	[This.ID]
145	0x0091	End.DX2 with PSD	[This.ID]
146	0x0092	End.DX2V with PSD	[This.ID]
147	0x0093	End.DT2U with PSD	[This.ID]
148	0x0094	End.DT2M with PSD	[This.ID]

Table 1: IETF - SRv6 Endpoint Behaviors

6. 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].

7. References

7.1. Normative References

[I-D.ietf-rtgwg-segment-routing-ti-lfa] Litkowski, S., Bashandy, A., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-11, 30 June 2023, <https://www.ietf.org/archive/id/draft-ietf-rtgwg-segment-routing-ti-lfa-11.txt>.

[I-D.hu-spring-segment-routing-proxy-forwarding] Hu, Z., Chen, H., Yao, J., Bowers, C., Yongqing, and Yisong, "SR-TE Path Midpoint Restoration", Work in Progress, Internet-Draft, draft-hu-spring-segment-routing-proxy-forwarding-24, 21 August 2023, <https://www.ietf.org/archive/id/draft-hu-spring-segment-routing-proxy-forwarding-24.txt>.

[I-D.ietf-rtgwg-srv6-egress-protection] Hu, Z., Chen, H., Chen, H., Wu, P., Toy, M., Cao, C., He T., Liu, L., Liu, X., "SRv6 Path Egress Protection", Work in Progress, Internet-Draft, draft-ietf-rtgwg-srv6-egress-protection-14, 29 August 2023, https://www.ietf.org/archive/id/draft-ietf-rtgwg-srv6-egress-protection-14.txt>

[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>.

[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>.

[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>.

7.2. Informative References

TBD

8. Acknowledgments

The authors would like to thank the following for their valuable contributions of this document:

Yisong Liu
China Mobile

Authors' Addresses

Weiqiang Cheng
China Mobile

Email: chengweiqiang@chinamobile.com

Wenying Jiang
China Mobile

Email: jiangwenying@chinamobile.com

Changwang Lin
New H3C Technologies

Email: linchangwang.04414@h3c.com

Zhibo Hu
Huawei Technologies

Email: huzhibo@huawei.com

Yuanxiang Qiu
New H3C Technologies

Email: qiuyuanxiang@h3c.com