Routing area
Internet-Draft
Intended status: Standards Track
Expires: May 19, 2021

S. Hegde
W. Lin
Juniper Networks Inc.
P. Shaofu
ZTE Corporation

November 15, 2020

Egress Protection for Segment Routing (SR) networks
draft-hegde-rtgwg-egress-protection-sr-networks-01

Abstract

This document specifies a Fast Reroute(FRR) mechanism for protecting IP/MPLS services that use Segment Routing (SR) paths for transport against egress node and egress link failures. The mechanism is based on egress protection framework described in [RFC8679]. The egress protection mechanism can be further simplified in Segment Routing networks with anycast SIDs and anycast Locators. This document addresses all kinds of networks that use Segment Routing transport such as SR-MPLS over IPv4, SR-MPLS over IPv6, SRv6 and SRm6.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

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 May 19, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 2
2. Egress Node Protection 3
  2.1. SR-MPLS Networks 4
  2.2. SRm6 Networks 5
  2.3. SRv6 Networks 5
3. Egress Link Protection 6
4. Security Considerations 7
5. IANA Considerations 7
6. Acknowledgments 7
7. References 7
  7.1. Normative References 7
  7.2. Informative References 8
Authors' Addresses 9

1. Introduction

Segment Routing Architecture as defined in [RFC8402] provides a simple and scalable MPLS control plane that removes state from transit nodes in the network. SRm6 as defined in [I-D.bonica-spring-sr-mapped-six] and SRv6 as defined in [I-D.ietf-spring-srv6-network-programming] provide Segment Routing transport in pure IPv6 networks where MPLS data plane is not used. End-to-End resiliency is very important to satisfy Service Level Agreements (SLA) such as 50ms convergence. The transport resiliency and fast rerouting are described in[I-D.ietf-rtgwg-segment-routing-ti-lfa] and [I-D.ietf-spring-segment-protection-sr-te-paths]. Egress node and egress link failures are not covered by these protection mechanisms. Egress node and link failures need to address moving the services to other nodes where the customer services are multi-homed. In traditional MPLS networks service labels (ex: L3VPN) are assigned

dynamically. The protector nodes need to learn the service labels advertised by primary nodes and build a local context table corresponding to each primary node that a node protects. This requires building local context tables and also specialized context table lookups as described in [RFC8679].

Egress protection can be simplified by statically assigning service labels on egress nodes. When a service is multi-homed to two or more egress nodes, the same service label can be assigned to the service on each egress node. This mechanism, coupled with using anycast SIDs for loopbacks, greatly simplifies the egress protection. Following the principles described in [RFC8679], this document specifies procedures which can be used to greatly simplify the operation of egress protection in a segment routing network. Egress protection for Multicast services is for FFS.

2. Egress Node Protection

```
                      services 1, ..., N
     ====================================> tunnel

   I ------ R1 ------- PLR --------------- E ----
ingress          penultimate-hop        egress    \
                       |  .           (primary    \
                       |  .            service     \
                       |  .            instances)   \
                       |  .                          \
                       |  .                           \   service
                       |  .                             destinations
                       |  .                           / (CEs, sites)
                       |  .                          /
                       |  . TI-LFA backup           /
                       |  . path                   /
                       |  .                       /
                       |  ...............        /
                       R2 --------------- P ----
                                      protector
                                     (protection
                                      service
                                      instances)
```

Figure 1: Reference topology

The reference topology from [RFC8679] has been reproduced in Figure 1 for ease of reading. The current document also uses the terminology defined in [RFC8679]. In the topology in Figure 1, service destinations are attached to two egress nodes. The two egress nodes

could be used in primary/protection mode or they could also be used in ECMP mode. When one of egress nodes fails, traffic should be switched to the other egress node and the convergence should be on the order of 50ms. The transport network is based on Segment Routing technology and could be using any of the SR-MPLS over IPv4, SR-MPLS over IPV6 , SRm6 or SRv6. The sections below describe the solution for each of the transport mechanisms.

2.1. SR-MPLS Networks

[RFC8402] describes the concept of anycast SIDs. Applying anycast SIDs to egress protection, the same IP address is configured as a loopback address on multiple egress nodes, and the same SID is advertised for this IP address. In the reference topology in Figure 1, E and P are associated with anycast loopback and corresponding anycast SID. The egress protected tunnel is considered logically destined to this anycast address and the egress protected tunnel always carries this anycast SID corresponding to the destination anycast address as the last SID. The egress protected service is hosted on both E and P. The egress protected tunnel can be used in primary/protector mode in which case, the anycast loopback MUST be advertised with better metric and the protector MUST advertise with an inferior metric. An egress protected service MUST advertise same service label from both E and P. The service label is assigned from the SRLB as defined in [RFC8402]. The egress node pairs that serve egress protected service MUST be able to allocate the same service label and hence MUST have overlapping local label space (SRLB) reserved for static assignment.

The TI-LFA procedures described in [I-D.ietf-rtgwg-segment-routing-ti-lfa] apply to the anycast prefixes. Based on the transport IGP topology, TI-LFA backup path is computed and programmed into the forwarding plane on the PLR nodes. The PLR SHOULD be configured to provide node protection for the failure. On the egress node E's failure, traffic on the PLR SHOULD be switched to the other egress node which is P. As the service label carried in the packet is understood by P as well, P will correctly send the traffic to the service destination.

Note that the micro-loop avoidance procedures as described in [I-D.bashandy-rtgwg-segment-routing-uloop] are applicable to anycast prefixes as well. When the anycast prefix is impacted by the failure event, a micro-loop avoiding path for the anycast prefix/anycast-SID will be programed during convergence. This mechanims does not affect the egress protection procedures described in this document.

The above procedure is applicable to SR-MPLS over IPv4 as well as SR-MPLS over IPv6 networks. In case of IPv4 networks, the anycast SIDs

are assigned to IPv4 loopbacks and in case of IPv6 networks, the anycast SIDs are assigned to IPv6 loopback addresses. The egress protected IP/MPLS services advertise the service prefix with the anycast address information in the message. In case of BGP based services such as L3vpn [RFC2547], the nexthop attribute carries the anycast address which is the logical tunnel destination address. On the ingress, when the service prefix is received, the service is mapped to the corresponding egress protected tunnel.

If some services are multi-homed to a different node for example, in the reference topology, if a service is multi-homed to E and another node P', then there SHOULD be another anycast address representing {E,P'}. The number of anycast loopbacks on a given node will be equal to the number of such {primary, protector} pairs a node belongs to. The egress protected service prefixes MUST carry the anycast address corresponding to the {primary, protector} pair in their next hop attribute.

When there is a single homed CE connected to the egress node, it SHOULD use a node loopback in the next hop attribute and should not use anycast loopback address.

2.2. SRm6 Networks

[I-D.bonica-spring-sr-mapped-six] defines segment routing applied to IPv6 networks which is optimized for high data rate forwarding. SRm6 control plane is very similar to SR-MPLS but it uses the IPv6 data plane. The egress node protection procedures described for SR-MPLS are applicable to SRm6 as well. Anycast loopback addresses are advertised and corresponding anycast SIDs are associated with the anycast addresses. The anycast SIDs in case of SRm6 are globally unique indices of size 16 or 32 bits.

The VPN services that require a label to identify the service are advertised as described in [I-D.ssangli-idr-bgp-vpn-srv6-plus]. The same PPSI value MUST be allocated to the service prefix on both the egress nodes on which the service is multi-homed. The TI-LFA procedures explained in Section 2.1 are applicable to SRm6 as well. After the CRH header is removed at the egress node, lookup is done based on PPSI which points to the correct service instance. Since the same PPSI is assigned on both nodes, the context table as described in [RFC8679] is not required to be built.

2.3. SRv6 Networks

[I-D.ietf-spring-srv6-network-programming] describes various types of SIDs used in SRv6 networks. The routing in the transport is based on the locators. Locators are most significant bits of the SID. In

order to achieve the egress protection functionality, anycast locators MUST be assigned on the egress nodes {E,P} where the service are multi-homed. The service SIDs MUST be derived from the anycast SID. The multi-homed service MUST be assigned with same service SID on both the egress nodes. It is recommended to provide mechanisms to statically configure the service SIDs which can easily serve the purpose of synchronized SID allocation on both nodes. As explained in the Section 2.1, if an egress node has services which are multi-homed to different nodes, then each such pair of node will need a separate locator assigned.

When there is a single homed CE connected to an egress, it MUST use a node specific locator to advertise service SID. It should not use service SIDs based on anycast locator

The TI-LFA procedure is applicable to anycast locators and each transport node in the transport IGP, installs a primary and backup path to the anycast locator. However, only the directly connected upstream PLR of the primary egress node will respond to the failure of the primary egress node and switch to the TI-LFA backup path. It is assumed that PLR has a backup path to alternate egress node which does not go via the primary egress node.

3. Egress Link Protection

```
                                                        Anycast loopback:
10.10.10.10
                                                        Anycast SID: 10 SRGB
(1000-2000)
                                                        Node loopback: 2.2.2.2

                           ---------- R1 ----------- PE2 -
                          /          (PLR)          (PLR)  \
    (          )         /            |               |     (          )
    (          )        /             |               |     (          )
    (  site 1  )-- PE1 |              |               R3    (  site 2  )
    (          )        \             |               |     (          )
    (          )         \            |               |     (          )
                          \           |               |    /
                           ---------- R2 ----------- PE3 -
                                                  (protector)
                                                  Anycast loopback:10.10.10.10
                                                  Anycast SID: 10 SRGB
(1000-2000)
                                                  Node loopback: 3.3.3.3
```

Figure 2: Egress Link Protection

The link from egress node towards the CE (service destination) fails, that failure needs to be protected. Egress link protection can be

achieved using similar means as egress node protection. section 6.2.2 of [I-D.ietf-rtgwg-bgp-pic] describes the procedures for protecting egress link failures in detail. When anycast ip address is used as BGP protocol nexthop, some additional considerations are required along with the procedures described in [I-D.ietf-rtgwg-bgp-pic]. In egress link failure case, egress node is the PLR and it has learned the service prefix from the other egress node. PLR egress node pre-establishes backup path to the other egress node and programs the forwarding plane with backup path. As the BGP based service prefixes advertise the anycast loopback address in the next hop attribute, the egress nodes will ignore the advertisement from other egress node. For Example, in the above Figure 2, when PE3 advertises a service prefix for site 2 with a next hop attribute of anycast loopback address, PE2 does not consider this advertisement and program a backup path towards PE3. To solve this problem, The egress nodes could advertise service prefixes with NEXT_HOP [RFC4271]attribute carrying anycast loopback as well as node specific loopback with a different RD [RFC2547].

4. Security Considerations

This document does not introduce any new security risks. For deploying this solution, security considerations described in [RFC8402], [I-D.bonica-spring-sr-mapped-six], [I-D.ietf-spring-srv6-network-programming] and [RFC8679] are applicable.

5. IANA Considerations

This document does not introduce any new IANA requests.

6. Acknowledgments

Thanks to Krzysztof Szarkowicz,Louis Chan and Chris Bowers for careful review and inputs.

7. References

7.1. Normative References

[I-D.bonica-spring-sr-mapped-six]
Bonica, R., Hegde, S., Kamite, Y., Alston, A., Henriques, D., Jalil, L., Halpern, J., Linkova, J., and G. Chen, "Segment Routing Mapped To IPv6 (SRm6)", draft-bonica-spring-sr-mapped-six-02 (work in progress), October 2020.

[I-D.ietf-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-ietf-spring-srv6-network-programming-24 (work in progress), October 2020.

[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>.

[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>.

7.2. Informative References

[I-D.bashandy-rtgwg-segment-routing-uloop]
Bashandy, A., Filsfils, C., Litkowski, S., Decraene, B., Francois, P., and P. Psenak, "Loop avoidance using Segment Routing", draft-bashandy-rtgwg-segment-routing-uloop-09 (work in progress), June 2020.

[I-D.ietf-rtgwg-bgp-pic]
Bashandy, A., Filsfils, C., and P. Mohapatra, "BGP Prefix Independent Convergence", draft-ietf-rtgwg-bgp-pic-12 (work in progress), August 2020.

[I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B., Francois, P., Voyer, D., Clad, F., and P. Camarillo, "Topology Independent Fast Reroute using Segment Routing", draft-ietf-rtgwg-segment-routing-ti-lfa-04 (work in progress), August 2020.

[I-D.ietf-spring-segment-protection-sr-te-paths]
Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu, "Segment Protection for SR-TE Paths", draft-ietf-spring-segment-protection-sr-te-paths-00 (work in progress), September 2020.

[I-D.ssangli-idr-bgp-vpn-srv6-plus]
Ramachandra, S. and R. Bonica, "BGP based Virtual Private Network (VPN) Services over SRv6+ enabled IPv6 networks", draft-ssangli-idr-bgp-vpn-srv6-plus-02 (work in progress), July 2019.

[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>.

[RFC2547] Rosen, E. and Y. Rekhter, "BGP/MPLS VPNs", RFC 2547, DOI 10.17487/RFC2547, March 1999, <https://www.rfc-editor.org/info/rfc2547>.

[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>.

[RFC7911] Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", RFC 7911, DOI 10.17487/RFC7911, July 2016, <https://www.rfc-editor.org/info/rfc7911>.

Authors' Addresses

Shraddha Hegde
Juniper Networks Inc.
Exora Business Park
Bangalore, KA 560103
India

Email: shraddha@juniper.net

Wen Lin
Juniper Networks Inc.

Email: wlin@juniper.net

Peng Shaofu
ZTE Corporation

Email: peng.shaofu@zte.com.cn