PIM Working Group
Internet Draft
Intended status: Informational
Expires: September 4, 2022

Yisong Liu
China Mobile
M. McBride
Futurewei
Z. Zhang
ZTE
J. Xie
Huawei
C. Lin
New H3C Technologies
Mar 4, 2022

Multicast-only Fast Reroute Based on Topology Independent Loop-free Alternate Fast Reroute
draft-liu-pim-mofrr-tilfa-05

Abstract

Multicast-only Fast Reroute (MoFRR) has been defined in [RFC7431], but the selection of the secondary multicast next hop only according to the loop-free alternate fast reroute, which has restrictions in multicast deployments. This document describes a mechanism for Multicast-only Fast Reroute by using Topology Independent Loop-free Alternate fast reroute, which is independent of network topology and can achieve covering more network environments.

Table of Contents

1. Introduction

As the deployment of video services, operators are paying more and more attention to solutions that minimize the service disruption due to faults in the IP network carrying the packets for these services. Multicast-only Fast Reroute (MoFRR) has been defined in [RFC7431], which can minimize multicast packet loss in a network when node or link failures occur by making simple enhancements to multicast routing protocols such as Protocol Independent Multicast (PIM). But the selection of the secondary multicast next hop only according to the loop-free alternate fast reroute in [RFC7431], and there are limitations in multicast deployments for this mechanism. This

document describes a new mechanism for Multicast-only Fast Reroute using Topology Independent Loop-free Alternate (TILFA) fast reroute, which is independent of network topology and can achieve covering more network environments.

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.

1.2. Terminology

This document use the terms defined in [RFC7431], and also use the concepts defined in [RFC7490]. The specific content of each term is not described in this document.

2. Problem Statement

In [RFC7431] section 3, the secondary Upstream Multicast Hop (UMH) of PIM for MoFRR is a loop-free alternate (LFA). However, the traditional LFA mechanism needs to satisfy at least one neighbor whose next hop to the destination node is an acyclic next hop, existing limitations in network deployments, and can only cover part of the network topology environments. In some network topology, the corresponding secondary UMH cannot be calculated, so PIM cannot establish a standby multicast tree and cannot implement MoFRR protection. Therefore, the current MoFRR of PIM is only available in the network topology applicable to LFA.

The remote loop-free alternate (RLFA) defined in [RFC7490] is extended from the LFA and can cover more network deployment scenarios through the tunnel as an alternate path. The RLFA mechanism needs to satisfy at least one node assumed to be N in the network that the fault node is neither on the path from the source node to the N node, nor on the path from the N node to the destination node. RLFA only has enhancement compared to LFA but still has limitations in network deployments.

[I-D.ietf-rtgwg-segment-routing-ti-lfa] defined a unicast FRR solution based on the TILFA mechanism. The TILFA mechanism can express the backup path with an explicit path, and has no constraint on the topology, providing a more reliable FRR mechanism. The unicast traffic can be forwarded according to the explicit path list as an alternate path to implement unicast traffic protection, and can achieve full coverage of various networking environments.

The alternate path provided by the TILFA mechanism is actually a Segment List, including one or more Adjacency SIDs of one or more links between the P space and the Q space, and the NodeSID of P space node. PIM can look up the corresponding node IP address in the unicast route according to the NodeSID, and the IP addresses of the two endpoints of the corresponding link in the unicast route according to the Adjacency SIDs, but the multicast protocol packets cannot be directly sent along the path of the Segment List.

PIM join message need to be sent hop-by-hop to establish a standby multicast tree. However, not all of the nodes and links on the unicast alternate path are included in the Segment List. If the PIM protocol packets are transmitted only in unicast mode, then equivalently the PIM packets are transmitted through the unicast tunnel like unicast traffic, and cannot pass through the intermediate nodes of the tunnel. The intermediate nodes of the alternate path cannot forward multicast traffic because there are no PIM state entries on the nodes. PIM needs to create entries on the device hop-by-hop and generate an incoming interface and an outgoing interface list. So it can form an end-to-end complete multicast tree for forwarding multicast traffic. Therefore, it is not possible to send PIM packets like unicast traffic according to the Segment List path and can only establish a standby multicast tree.

3. Solution

A secondary Upstream Multicast Hop (UMH) is a candidate next-hop that can be used to reach the root of the tree. In This document the secondary UMH is based on unicast routing to find the Segment List calculated by TILFA.

It is available in principle that the path information of the Segment List is added to the PIM packets to guide the hop-by-hop RPF selection. The IP address of the node corresponding to the NodeSID can be used as the segmented root node, and the IP addresses of the interfaces at both endpoints of the link corresponding to the Adjacency SID can be used directly as the local upstream interface and upstream neighbor.

For the PIM protocol, the PIM RPF Vector attribute was defined in [RFC5496], which can carry the node IP address corresponding to the NodeSID. The explicit RPF Vector was defined in [RFC7891], which can carry the peer IP address corresponding to the Adjacency SID.

This document can use the above two RPF Vector standards and does not need to extend the PIM protocol, to establish the standby multicast tree according to the Segment List calculated by TILFA,

and can achieve full coverage of various networking environments for MoFRR protection of multicast services.

Assume that the Segment List calculated by TILFA is (NodeSID(A), AdjSID(A-B)). Node A belongs to the P Space, and node B belongs to the Q space. The IP address corresponding to NodeSID(A) can be looked up in the local link state database of the IGP protocol, and can be assumed to be IP-a. The IP addresses of the two endpoints of the link corresponding to AdjSID(A-B) can also be looked up in the local link state database of the IGP protocol, and can be assumed to be IP-La and IP-Lb.

In the procedure of PIM, IP-a can be looked as the normal RPF vector attribute and added to the PIM join packet. IP-La can be looked as the local address of the incoming interface, and IP-Lb can be looked as the address of the upstream neighbor. So IP-Lb can be added to the PIM join packet as the explicit RPF Vector attribute.

The PIM protocol firstly can select the RPF incoming interface and upstream towards IP-a, and can join hop-by-hop to establish the PIM standby multicast tree until the node A. On the node A, IP-Lb can be looked as the PIM upstream neighbor. The node A can find the incoming interface in the unicast routing table according to the IP-Lb and IP-Lb is used as the RPF upstream address of the PIM join packet to the node B.

After the PIM join packet is received on the node B, the PIM protocol can find no more RPF Vector attributes and select the RPF incoming interface and upstream towards the multicast source directly, and then can continue to join hop-by-hop to establish the PIM standby multicast tree until the router directly connected the source.

4. Illustration

This section provides an illustration of MoFRR based on TI-LFA. The example topology is shown in Figure 1. The metric of R3-R4 link is 100, and the metrics of other links are 10. All the link metrics are bidirectional.

```
        <-----Priamry Path--- (S,G) Join

[S]---(R1)---(R2)******(R6)-------[R]
               |        |
        <---   |        |   |
           |   |        |   |
           |   |       (R5) |
           |   |        |   |
```

```
           |   |        |   |
           |   |        |   |
           | (R3)------(R4) |
           |                |
           --Secondary Path--
```

Figure 1: Sample Topology

The IP addresses and MPLS SIDs, which may be involved in the MoFRR calculation, are configured as following:

```
Node    IP Address   Node SID
R4      4.4.4.4/32   16004

Link    IP Address   Adjacency SID
R3->R4  14.0.0.1/24  24001
R4->R3  14.0.0.2/24  24002
```

The primary path of the PIM join packet is R6->R2->R1, and the secondary path is R6->R5->R4->R3->R2->R1. However, the traditional LFA does not work properly for the secondary path, because the shortest path to R2 from R5 (or from R4) still goes through R6-R2 link. In this case, R6 needs to calculate the secondary UMH using the proposed MoFRR method based on TI-LFA.

According to the TI-LFA algorithm, P-Space and Q-Space are shown in Figure 2. The TI-LFA repair path consists of the Node SID of R4 and the Adjacency SID of R4->R3. The repair segment list is (16004, 24002).

```
           ........
           .      .
[S]---(R1)---(R2)******(R6)---[R]
           .   |  .     |
           .   |  .  +++|++++
           .   |  .  +  |   +
           .   |  .  + (R5) +
           .   |  .  +  |   +
           .   |  .  +  |   +
           .   |  .  +  |   +
           . (R3)------(R4) +
           .      .  +      +
           ........  ++++++++
           Q-Space    P-Space
```

Figure 2: P-Space and Q-Space

In the procedure of PIM, the IP addresses associated with the repair segment list are looked up in the IGP link state database. The Node SID 16004 corresponds to 4.4.4.4, which will be carried in the RPF Vector Attribute. The Adjacency SID 24002 corresponds to local address 14.0.0.2 and remote peer address 14.0.0.1, and 14.0.0.1 will be carried in the Explicit RPF Vector Attribute. Therefore, R6 installs the secondary UMH with these RPF Vectors.

The forwarding of PIM join packet along the secondary path is shown in Figure 3.

```
   +--------+
   |Type = 0|
   |4.4.4.4 |
   +--------+     +--------+
   |Type = 4|     |Type = 4|
   |14.0.0.1|     |14.0.0.1|
   +--------+     +--------+    No RPF Vector

R6----->R5---->R4------------>R3----->R2---->R1
```

Figure 3: Forwarding PIM Join Packet along Secondary Path

R6 inserts two RPF Vector Attributes in the PIM join packet, which are 4.4.4.4 of Type 0 (RPF Vector Attribute) and 14.0.0.1 of Type 4 (Explicit RPF Vector Attribute). Then R6 forwards the packet along the secondary path.

When R5 receives the packet, R5 performs a unicast route lookup of the first RPF Vector 4.4.4.4 and sends the packet to R4.

R4 is the owner of 4.4.4.4, so it removes the first RPF Vector from the packet and forwards according to the following RPF Vector. R4 sends the packet to R3 according to the next RPF Vector 14.0.0.1, since its PIM neighbor R3 corresponds to 14.0.0.1.

When R3 receives the packet, as the owner of 14.0.0.1, it removes the RPF Vector. Then the packet has no RPF Vector, and will be forwarded to the source through R3->R2->R1 according to unicast routes.

After the PIM join packet reaches R1, a secondary multicast tree, R1->R2->R3->R4->R5->R6, is established hop-by-hop for (S, G) by MoFRR based on TI-LFA.

The above procedures can also work in IPv6 data plane. The TI-LFA path computation algorithm in the SRv6 data plane is the same as in the SR-MPLS data plane. Instead of MPLS labels, SRv6 SIDs are used

to build repair list. Similarly, the RPF Vector Attributes and the Explicit RPF Vector Attributes will contain IPv6 addresses instead of IPv4 addresses.

5. IANA Considerations

No IANA actions are required for this document.

6. Security Considerations

This document does not change the security properties of PIM. For general PIM-SM protocol Security Considerations, see [RFC7761].

7. References

7.1. Normative References

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

[RFC5384] Boers, A., Wijnands, I., and E. Rosen, "The Protocol Independent Multicast (PIM) Join Attribute Format", RFC 5384, November 2008

[RFC5496] Wijnands, IJ., Boers, A., and E. Rosen, "The Reverse Path Forwarding (RPF) Vector TLV", RFC 5496, March 2009

[RFC7431] Karan, A., Filsfils, C., Wijnands, IJ., Ed., and B. Decraene, "Multicast-Only Fast Reroute", RFC 7431, August 2015

[RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, April 2015

[RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I.,Parekh, R., Zhang, Z., and L. Zheng, "Protocol IndependentMulticast - Sparse Mode (PIM-SM): Protocol Specification(Revised)", RFC 7761, March 2016

[RFC7891] Asghar, J., Wijnands, IJ., Ed., Krishnaswamy, S., Karan, A., and V. Arya, "Explicit Reverse Path Forwarding (RPF) Vector", RFC 7891, June 2016

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

[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", draft-ietf-rtgwg-segment-routing-ti-lfa-08, work-in-progress, January 2022

7.2. Informative References

TBD

8. Acknowledgments

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

TBD

Authors' Addresses

Yisong Liu
China Mobile
China
Email: liuyisong@chinamobile.com

Mike McBride
Futurewei Inc.
USA
Email: michael.mcbride@futurewei.com

Zheng(Sandy) Zhang
ZTE Corporation
China
Email: zzhang_ietf@hotmail.com

Jingrong Xie
Huawei Technologies
China
Email: xiejingrong@huawei.com

Changwang Lin
New H3C Technologies
China
Email: linchangwang.04414@h3c.com