PIM Working Group
Internet Draft
Intended status: Standards Track
Expires: May 23, 2020

Yisong Liu
Individual
M. McBride
Futurewei

Nov 23, 2019

Multicast-only Fast Reroute Based on Topology Independent Loop-free Alternate Fast Reroute
draft-liu-pim-mofrr-tilfa-01

Status of this Memo

Copyright Notice

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) and Multipoint LDP (mLDP). 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 and mLDP 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 and MLDP cannot establish a standby multicast tree and cannot implement MoFRR protection. Therefore, the current MoFRR of PIM and MLDP 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 and MLDP 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.

Both the PIM join message and the MLDP Label Mapping 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 and MLDP protocol packets are transmitted only in unicast mode, then equivalently the PIM and MLDP 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 or MLDP 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. MLDP needs to create entries on the device hop-by-hop and generate an incoming label and an outgoing label list. So it can form an end-to-end complete multicast tree for forwarding multicast traffic. Therefore, it is not possible to send PIM and MLDP packets like unicast traffic according to the Segment List path and establish a standby multicast tree.

It is available in principle that the path information of the Segment List is added to the PIM and MLDP 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, but there is currently no field in protocol packet to carry the explicit path specified by the Segment List. 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, but if there are multiple same peer IP addresses corresponding to the Adjacency SID (i.e. anycast IP address), the upstream neighbor of RPF selection may be different from the actual upstream link corresponding to the Adjacency SID, which can make the PIM join path and the TILFA calculation path inconsistent. For the MLDP protocol, there is also no field defined in the MLDP protocol Label Mapping message that can carry the explicit path of the Segment List.

3. Solution

An 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 Segment List calculated by TILFA. With MoFRR, The procedures for determining the secondary UMH and establishing standby multicast tree are different for PIM and mLDP.

This document extends the PIM and mLDP 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.

3.1. Secondary UMH Selection in PIM

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 and IP-Lb can be looked as the RPF Vector attribute of the adjacency relationship, called Adjacency RPF Vector, which is a new type of PIM join attributes, and added to the PIM join packet too.

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 one PIM neighbor. If there are multiple PIM neighbors with the same address IP-Lb, all of the corresponding local interfaces on the node A need to be checked. The interface that is the only one with the IP address IP-La can be looked as the RPF incoming interface. The node A can send the PIM join packet to the node B on the interface of IP-La, and IP-Lb is used as the RPF upstream address of the PIM join.

After the PIM join packet is received on the node B, the PIM protocol can find no more join 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.

3.2. Secondary UMH Selection in MLDP

In the procedure of MLDP, Explicit path TLV is newly defined in MLDP Label Mapping message to carry IP-a, IP-La and IP-Lb, which is contained in the field of Optional Parameters. IP-a can be looked as the segmented root node address and is added as the Node Address Sub TLV in the Explicit path TLV. IP-La and IP-Lb are added as the Adjacency Address Sub TLV in the Explicit Path TLV.

The MLDP protocol can look up the upstream interface and the upstream LSR in the unicast route to IP-a, and can send the Label Mapping message hop-by-hop to establish the standby MPLS multicast tree to the node A. After the message is received on the node A, the Node Address Sub TLV corresponding to the IP-a can be deleted from the Label Mapping message.

On the node A IP-Lb can be looked as one MLDP neighbor. If there are multiple MLDP neighbors with the same address IP-Lb, all of the corresponding local interfaces on the node A need to be checked. The interface that is the only one with the IP address IP-La can be looked as the upstream interface. The node A can send MLDP Label mapping message to the node B, and IP-Lb is used as the upstream LSR address.

After the message is received on the node B, the Adjacency Address Sub TLV corresponding to the IP-La and IP-Lb is deleted from the Label Mapping message and if there is no more any sub TLV in the Explicit Path TLV then the TLV should be deleted. The MLDP protocol can select the upstream interface and the upstream LSR in the unicast route to the original root node directly, and can continue to send the Label Mapping message to establish the standby MPLS multicast tree to the original root node.

3.3. Extension Protocol Fields Conflict

PIM Adjacency RPF Vector attribute is newly defined in join attributes. If there are conflicts from multiple downstream PIM neighbors, the mechanism in [RFC5384] Section 3.3.3 can be used to select a PIM downstream neighbor with a numerically smallest IP address. If at least two neighbors have the same IP address, the interface index MUST be used as a tie breaker.

In the Explicit Path TLV newly defined in MLDP Label Mapping message, if there are conflicts from multiple downstream MLDP neighbors, including the inconsistency of the Sub TLV types, and the inconsistency of the Sub TLV contents, and the inconsistency of the Sub TLV sequences, it is also recommended to use the mechanism in [RFC5384] Section 3.3.3.

4. Packet Format

This section describes the format of PIM and mLDP protocol packet extension introduced by this document.

4.1. PIM Join Message Extension

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Addr Family   | Encoding Type | Rsrvd   |S|W|R|  Mask Len     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Source Address
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....
|F|E| Attr_Type | Length        | Value
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....
|F|E| Attr_Type | Length        | Value
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....
     .                     .                      .
     .                     .                      .
```

The original PIM join attribute already has been defined in [RFC5384]

Attr_Type:

0- Vector ;

4- Explicit RPF Vector ;

Other existing definitions are not related to RPF Vector Attribute.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|E| Type      | Length        |        Value
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-.......
```

The definition of Adjacency RPF Vector attribute

F bit: 0, indicating that the unrecognized device does not forward the attribute

E bit: indicates the last join attribute

Type: TBD

Length: depends on the address family of Encoded-Unicast address, including the length of 2 addresses.

Value: Encoded-Unicast Address format defined in [RFC7761] Section 4.9.1, including 2 addresses. The first one indicates the address of the local interface, and the second one indicates the address of the peer interface. Only the case of the same address family is supported.

4.2. MLDP Label Mapping Message Extension

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   Label Mapping (0x0400)    |      Message Length           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Message ID                                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     FEC TLV                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Label TLV                                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Optional Parameters                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The LDP Label Mapping message format is defined in [RFC5036] Section 3.5.7. The MLDP P2MP protocol uses the message to establish a P2MP multicast tree. The Optional Parameters field can be extended to carry the node or link IP address list specified by the Segment List.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |U|F|        Type               |            Length             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    |                             Value                             |
    ~                                                               ~
    |                                                               |
    |                               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The TLV format definition in [RFC5036] Section 3.3 can be used for the Explicit Path TLV carrying the specified path of the Segment List.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |1|0|        TBD1               |            Length             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |               Node Address Sub TLV                            |
   |               Adjacency Address Sub TLV                       |
   ~                                                               ~
   |                                                               |
   |                               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The definition of Explicit Path TLV:

U bit: Unknown TLV bit. 1 indicates the unknown TLV MUST be silently ignored and the rest of the message processed as if the unknown TLV did not exist.

F bit: Forward unknown TLV bit. 0 indicates the unknown TLV is not forwarded with the containing message.

Type: TBD1

Length: contains all Sub-TLV lengths

Value: Contains one or more Sub-TLVs, which are recorded in the order of TILFA's Segment List. There are two types of Sub TLVs now. One of the two types is called Node Address Sub TLV which carries the node IP address corresponding to the NodeSID, and the other is called Adjacency Address Sub TLV which carries the local interface address and the peer interface address corresponding to the Adjacency SID.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |1|0|E|      Type = TBD2        |            Length             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 |                      Node Address                             |
 ~                                                               ~
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Node Address Sub TLV carrying the node IP address corresponding to the NodeSID

U bit: 1 indicates the unknown TLV MUST be silently ignored and the rest of the message processed as if the unknown TLV did not exist. F bit: 0 indicates the unknown TLV is not forwarded with the containing message.

E bit: 1 indicates the last Sub TLV.

Type: TBD2

Length: IPv4 address 4 octet, IPv6 address 16 octet

Value: The IP address of the node corresponding to the NodeSID in the Segment List generated by TILFA

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |1|0|E|      Type = TBD3         |             Length            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                                |
    |                 Local Interface Address                        |
    |                 Remote Interface Address                       |
    ~                                                                ~
    |                                                                |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Adjacency Address Sub TLV carrying the local interface address and the peer interface address corresponding to the Adjacency SID

U bit: 1 indicates the unknown TLV MUST be silently ignored and the rest of the message processed as if the unknown TLV did not exist.
F bit: 0 indicates the unknown TLV is not forwarded with the containing message.

E bit: 1 indicates the last Sub TLV.

Type: TBD3

Length: IPv4 address 8 octet, IPv6 address 32 octet

Value: The IP address of the local interface and the IP address of the peer interface corresponding to the Adjacency SID in the Segment List generated by TILFA must be recorded in order, and MUST be the same address family.

5. IANA Considerations

This document requests IANA to assign a registry for Adjacency RPF Vector in PIM Join Attribute and the Explicit Path TLV Node Address Sub TLV, Adjacency Address Sub TLV in the Optional Parameters field of MLDP P2MP Label Mapping message. The assignment is requested permanent for IANA when this document is published as an RFC. The string TBD, TBD1, TBD2 and TBD3 should all be replaced by the assigned values accordingly.

6. Security Considerations

For general PIM-SM protocol Security Considerations, see [RFC7761].

For general MLDP protocol Security Considerations, see [RFC6388]

TBD

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.

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, October 2007

[RFC5286] Atlas, A., Ed., and A. Zinin, Ed., "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, September 2008

[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

[RFC6388] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, November 2011

[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., 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-01 (Expired), March 2019

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
Individual

Email: liuyisong.ietf@gmail.com

Mike McBride
Futurewei Inc.

Email: michael.mcbride@futurewei.com