MBONED WG
Internet-Draft
Intended status: Informational
Expires: 5 January 2026

G. Shepherd
Cisco Systems, Inc.
Z. Zhang, Ed.
ZTE Corporation
Y. Liu
China Mobile
Y. Cheng
China Unicom
G. Mishra
Verizon Inc.
4 July 2025

Multicast Redundant Ingress Router Failover
draft-ietf-mboned-redundant-ingress-failover-07

Abstract

This document analyzes the redundant ingress router failover problem of a multicast domain, and analyzes the possible backup modes and advantages of each mode when deploying multiple ingress devices to forward the same multicast flow in a multicast domain.

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 5 January 2026.

Table of Contents

1. Introduction

Multicast redundant ingress router failover is an important issue in multicast deployments, especially in backbone multicast domains or multicast provider domains. Backbone multicast domains or multicast provider domains are referred to as multicast domains in the following sections. A multicast domain is a domain used to forward multicast flow based on specific multicast technologies, such as PIM [RFC7761], BIER [RFC8279], P2MP TE tunnel [RFC4875], MLDP [RFC6388], etc. Static configuration, tunnel based technologies, such as AMT [RFC7450], SR P2MP policies [I-D.ietf-pim-sr-p2mp-policy] can also be used. The domain may or may not be directly connected to the actual multicast source and receivers.

The ingress device of the multicast domain, such as the ingress router, can be connected to the multicast source by a single hop or multiple hops. In PIM, it is also called the first hop router, in BIER, it is called the BFIR, and in P2MP TE tunnel or MLDP, it is called the ingress LSR.

The egress device of the multicast domain, such as the egress router, may be connected to the multicast receiver by a single hop or multiple hops. In PIM, it is also called the last hop router, in BIER, it is called the BFER, and in P2MP TE tunnel or MLDP, it is called the egress LSR.

In order to ensure the reliability of multicast flow, there may be two or more ingress devices or egress devices in the multicast domain. That means the same multicast flow may enter the multicast domain from multiple ingress devices of the multicast domain. This draft does not discuss the protection method between the ingress device and the multicast source, between the egress device and the receiver, nor does it discuss the details of the technologies such as PIM and BIER. It only discusses the issue of failover of the ingress router of the multicast domain.

This document discusses the deployment of multiple ingress devices in a multicast domain. When a fault occurs, the switching method from the primary ingress device to the backup ingress device and the common fault detection methods are discussed. The advantages and disadvantages of the switching methods are analyzed to provide a reference for multicast deployment.

2. Terminology

The following abbreviations are used in this document:

IR: The ingress router for multicast flows in a multicast domain.

ER: The egress router for multicast flows in a multicast domain.

SIR: The IR responsible for sending the multicast flow, or the IR whose flow is received by the ER, is called Selected-IR, or SIR for short.

BIR: The IR may or may not send multicast flows. Multicast flows from IR will not be accepted by ER. Once SIR fails, IR will replace the role of SIR and multicast flows from IR will be accepted by ER. This IR is called backup IR, or BIR for short.

3. Multicast Redundant Ingress Router Failover

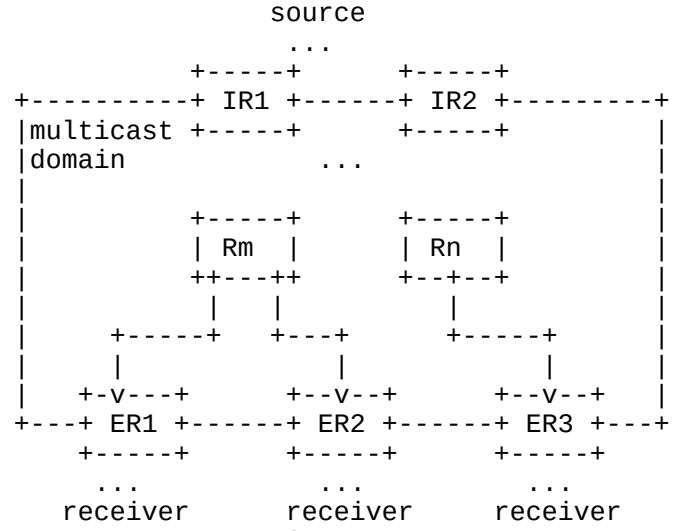

Figure 1

This is a common multicast networking scenario. The multicast domain includes the area from IR to ER. The flow sent by the multicast source enters the multicast domain from at least one IR, is forwarded in the multicast domain, reaches the ER, is forwarded by the ER, and finally the receiver receives the multicast flow.

The ingress device IR of the multicast domain is a key node for the normal forwarding of multicast flows. When two or more IRs are deployed, there may be multiple protection modes for IR, such as cold standby, warm standby and hot standby. These modes are also described in [RFC9026]. However, [RFC9026] mainly focuses on signaling notifications in MVPN scenarios and does not involve the protection mode of multiple ingress devices in the multicast domain and the impact on multicast flow transmission in the multicast domain.

As shown in Figure 1, a same multicast flow enters the multicast domain from two IRs. Both IRs are UMH (Upflow Multicast Hop) candidates of ER. Different multicast technologies may be used in the multicast domain according to the deployment of the network administrator. Assuming that PIM technology is used, two multicast trees can be pre-established with two IRs as roots.

3.1. Swichover

When a node or link in the multicast domain fails, the forwarding of multicast flow may be affected. However, it is not necessary to switch multicast flow from SIR to BIR in all cases. The following are situations where switching is not required:

* When PIM is used as the multicast forwarding protocol in a domain, a forwarding tree of (S, G) or (*, G) is pre-built. When a node other than SIR or a link in the forwarding tree fails, the tree is partially rebuilt.

* When BIER is used as the multicast forwarding protocol in a multicast domain, when a node other than SIR or a link in the domain fails, there is no need to rebuild the forwarding path, BIER forwarding will be restored as the IGP route converges.

* When P2MP TE tunnel or MLDP is used as the multicast forwarding protocol in a multicast domain, a forwarding LSP is pre-established. When a node other than the SIR in the LSP or a link in the domain fails, the LSP may be partially rebuilt.

* When a static multicast tree or SR P2MP policy is used in a multicast domain, when a node other than the SIR on the forwarding path or a link has a problem, the controller needs to recalculate a new forwarding path to bypass the faulty node or link.

When a critical failure occurs, it is necessary to switch from SIR to BIR, for example: SIR encounters a device failure, or the forwarding channel between SIR and ER fails, causing ER to be unable to receive multicast flows from SIR, and this failure cannot be restored in a short time. At this time, the multicast flow will be forwarded by BIR. ER receives the flow forwarded by BIR and forwards it to the receiver.

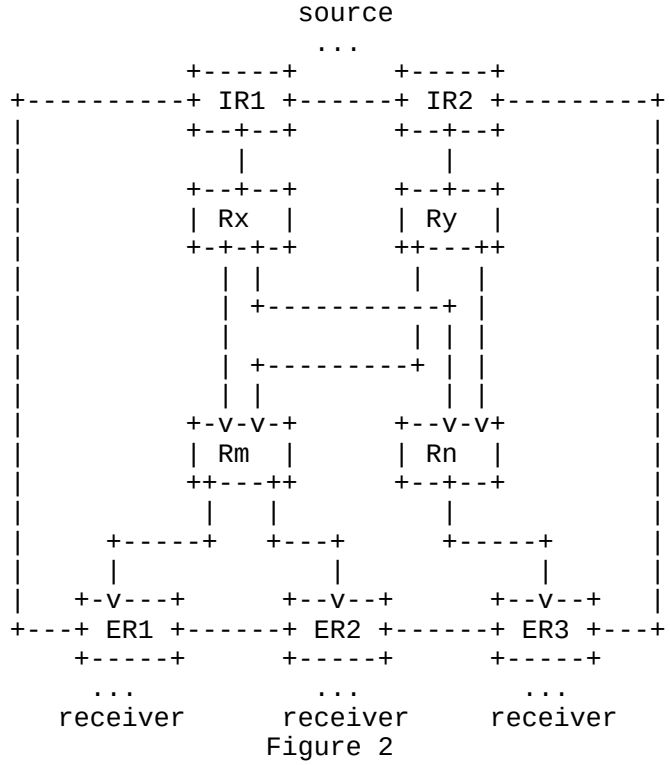

Figure 2

For example, in Figure 2, there is only one path in some areas of the network. IR1 and Rx are key nodes in the domain. When IR1 or Rx fails, there is no other path between IR1 and ER.

* When PIM is used in the multicast domain, Rm and Rn can select Ry as the upflow node, send Join messages, and build a new tree with IR2 as the root.

* When BIER is used in the multicast domain, IR2 should be responsible for the forwarding role and forward flow to ER.

* When P2MP TE tunnel or MLDP is used in the domain, LSP initiated from IR2 can be built and replace the LSP initiated from IR1.

* When a static multicast tree or SR P2MP policy is used in the multicast domain, the controller should build a new forwarding path with IR2 as the root to forward the multicast flow to ER.

3.2. Failure detection

The IR node itself and the key forwarding link between IR and ER are factors that affect traffic forwarding within the multicast domain.

In order to achieve fast switching, BIR can establish a forwarding channel with ER in advance and monitor the status of SIR. When the SIR node fails, it will take over the work of SIR. BIR can establish a BFD [RFC5880] session with SIR to detect the SIR status, or it can be detected by ping and other methods. However, it should be noted that the detection between BIR and SIR does not represent the actual forwarding path status between SIR and ER. When SIR is working normally, only the link between BIR and SIR fails, which may cause BIR to make wrong judgments and switch, thereby generating unnecessary duplicate flow. In this case, ER must support selective reception and be compatible with IR switching errors.

There may be problems with the forwarding path between SIR and ER, but the link between BIR and SIR is normal and cannot be detected by BIR. Therefore, ER can also detect the forwarding path between SIR and ER and actively switch to BIR to forward flow when problems are found. The detection between SIR and ER can be based on multipoint BFD [RFC8562]. When BIER is used to forward flow in the multicast domain, the detection between SIR and ER can also be based on BIER BFD [I-D.ietf-bier-bfd]. When MPLS is used to forward flow in the multicast domain, BFD [RFC5884] based on MPLS LSP can be used for detection.

4. Stand-by Modes

Detection and IR switching can be three modes: cold standby, warm standby, and hot standby. When the three modes are used to protect IR, the transmission mode of multicast flow in the multicast domain is different, and the impact on the network is also different.

When the multicast domain uses the PIM protocol to forward flow, ER will establish a multicast tree to BIR through signaling. When the multicast domain uses BIER to forward flow, ER will notify BIR the request to receive multicast flow through the BIER overlay protocol. When the multicast domain uses P2MP TE or MLDP to forward flow, a multicast forwarding channel is established from BIR to ER. The PIM multicast tree with BIR as the root and the P2MP TE or MLDP tunnel from BIR to ER can also be established in advance, and ER directly notifies BIR to use the multicast tree or tunnel for forwarding.

4.1. Cold Standby Mode

In cold standby mode, ER selects a SIR (e.g. IR1 in Figure 1) as the SIR and signals it to obtain the multicast flow.

When ER finds that it cannot receive the flow from IR1 through the detection means in Section 3.2, ER signals IR2 to obtain the multicast flow.

* For IR, IR (including SIR and BIR) only performs the normal operation of forwarding the flow according to ER request.

* For ER, ER must select an IR as the SIR and signal it. When the SIR fails or the path between SIR and ER fails, ER must signal BIR to obtain the flow.

* For intermediate routers, they know nothing about the role of IR, they only forward packets. There is no duplicate packets in the domain.

In this scenario, the BIR does not need to detect the status of the SIR. During the IR switching process, packet loss may occur because of the need for signaling interaction. Even if a PIM multicast tree or P2MP TE/MLDP tunnel is established in advance, packet loss may still occur.

4.2. Warm Standby Mode

In warm standby mode, the ER will signal to the SIR and BIR, such as IR1 and IR2 in Figure 2, that it needs to receive flow. The SIR (such as IR1) forwards the flow to the ER. The BIR (such as IR2) must not forward flow to the ER before the SIR fails. The BIR can detect the SIR status by the method described in Section 3.2, and automatically forward flow to the ER when the SIR fails.

* Normally, the SIR forwards flow to the ER. When the SIR fails or the path between the SIR and the ER fails, the BIR must start forwarding flow to the ER. The BIR can detect node failures in the SIR using the method described in Section 3.2, but may lack the method to detect path failures from the SIR to the ER.

* The ER does not distinguish between the SIR and the BIR. The ER only signals to both that it needs to receive a certain flow.

* For the intermediate routers, they do not know the difference between the IRs, and they are only responsible for packet forwarding. There are no duplicate packets in the domain.

When the BIR detects the SIR failure and starts forwarding flow, packet loss will occur during the switchover.

In some deployments, the SIR and BIR may be responsible for different multicast flows to share the load. For a certain multicast flow, the SIR may be IR1, and for another multicast flow, the SIR may be IR2. For example, IR1 sends some multicast flows to ERs and IR2 sends other multicast flows to ERs. Another possible deployment is that two IRs can be responsible for different ERs for the same multicast flow. If IR1 detects a failure between IR1 and ERs, IR1 may notify IR2 to forward flow to these ERs.

4.3. Hot Standby Mode

In hot standby mode, the ER signals both IRs that it wants to receive a certain flow. Both IRs send flows to the ER. The ER must discard duplicate flows from one of the IRs. In this case, there is no SIR or BIR. Only the ER knows which IR is the SIR.

* In this mode, the IR does not need to know the role of the SIR or BIR, IR only forwards the flow based on the request received from the ER.

* ER will send flow reception signals to both IRs and discard the duplicate flow from the backup BIR when it receives a duplicate flow. After switching the ER receives and forwards the flow from the BIR. It should be noted that the ER may choose different SIRs or BIRs for different multicast flows.

* Intermediate routers do not know the role of the IR, they only forward packets. There are duplicate packets within the domain.

In this mode, BIR does not need to detect the status of SIR. ER will detect the failure of SIR. Since duplicate flow packets arrive at ER, although packet loss may occur when ER switches to receive and forward flow from BIR, the packet loss is very small compared to the previous two modes.

4.4. Summary

The following table is a simple comparison of the three modes. "SIR failover" means that the SIR fails or the path between the SIR and the ER fails.

role	Cold Mode	Warm Mode	Hot Mode
IR	Forwards flow based on ER's request.	Acting as either SIR or BIR, BIR must not forward flow to ER until SIR fails over.	Does not need to know SIR or BIR role, just forwards flow based on ER's request.
ER	Must select an IR as SIR to signal request, signals BIR to request flow when SIR fails over.	Does not select SIR or BIR, just signals both of them.	Signals both SIR and BIR. Drops duplicate flow from BIR until SIR fails over.
Intermediate routers	Know nothing about SIR or BIR. Do not forward duplicate flow.	Know nothing about SIR or BIR. Do not forward duplicate flow.	No knowledge of SIR or BIR. Forward duplicate flow.

Table 1

Cold standby mode is the easiest to implement, but has the longest convergence time.

Warm standby mode has a moderate packet loss rate and convergence time, but it is difficult for BIR to know the path failure between SIR and ER.

Hot standby mode has the lowest packet loss rate, but there is duplicated packet forwarding within the domain, which consumes more bandwidth. For example, in the MVPN scenario, the hot root standby mode described in Section 5 [RFC9026] is the best recommended method for MVPN fast failover optimization. There may be duplicated packet forwarding within the domain, which will be discarded according to the provisions of [RFC9026] Section 6 and [RFC6513] Section 9.1.

For network administrators, the most appropriate standby mode should be selected based on the actual network deployment, such as whether there is enough bandwidth to accommodate duplicate flow.

5. IANA Considerations

This document does not have any requests for IANA allocation.

6. Security Considerations

This document adds no new security considerations.

7. References

7.1. Normative References

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <https://www.rfc-editor.org/info/rfc4875>.

[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, DOI 10.17487/RFC6388, November 2011, <https://www.rfc-editor.org/info/rfc6388>.

[RFC7450] Bumgardner, G., "Automatic Multicast Tunneling", RFC 7450, DOI 10.17487/RFC7450, February 2015, <https://www.rfc-editor.org/info/rfc7450>.

[RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March 2016, <https://www.rfc-editor.org/info/rfc7761>.

[RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <https://www.rfc-editor.org/info/rfc8279>.

7.2. Informative References

[I-D.ietf-bier-bfd]
Xiong, Q., Mirsky, G., hu, F., Liu, C., and G. S. Mishra, "BIER BFD", Work in Progress, Internet-Draft, draft-ietf-bier-bfd-08, 26 February 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-bier-bfd-08>.

[I-D.ietf-pim-sr-p2mp-policy]
Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. J. Zhang, "Segment Routing Point-to-Multipoint Policy", Work in Progress, Internet-Draft, draft-ietf-pim-sr-p2mp-policy-12, 23 May 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-pim-sr-p2mp-policy-12>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <https://www.rfc-editor.org/info/rfc5880>.

[RFC5884] Aggarwal, R., Kompella, K., Nadeau, T., and G. Swallow, "Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)", RFC 5884, DOI 10.17487/RFC5884, June 2010, <https://www.rfc-editor.org/info/rfc5884>.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC8562] Katz, D., Ward, D., Pallagatti, S., Ed., and G. Mirsky, Ed., "Bidirectional Forwarding Detection (BFD) for Multipoint Networks", RFC 8562, DOI 10.17487/RFC8562, April 2019, <https://www.rfc-editor.org/info/rfc8562>.

[RFC9026] Morin, T., Ed., Kebler, R., Ed., and G. Mirsky, Ed., "Multicast VPN Fast Upstream Failover", RFC 9026, DOI 10.17487/RFC9026, April 2021, <https://www.rfc-editor.org/info/rfc9026>.

Authors' Addresses

Greg Shepherd
Cisco Systems, Inc.
170 W. Tasman Dr.
San Jose,
United States of America
Email: gjshep@gmail.com

Zheng Zhang (editor)
ZTE Corporation
Nanjing
China
Email: zhang.zheng@zte.com.cn

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

Ying Cheng
China Unicom
Beijing
China
Email: chengying10@chinaunicom.cn

Gyan Mishra
Verizon Inc.
Email: gyan.s.mishra@verizon.com