

PIM Working Group
Internet-Draft
Intended status: Informational
Expires: September 11, 2020

H. Asaeda
NICT
LM. Contreras
Telefonica
March 10, 2020

Multiple Upstream Interface Support for IGMP/MLD Proxy
draft-asaeda-pim-multiif-igmpmldproxy-04

Abstract

This document describes the way of supporting multiple upstream interfaces for an IGMP/MLD proxy device. The proposed extension enables an IGMP/MLD proxy device to receive multicast sessions/channels through the different upstream interfaces. The upstream interface is selected based on the subscriber address prefixes, channel/session IDs, and interface priority values. A mechanism for upstream interface takeover that enables to switch from an inactive upstream interface to an active upstream interface is also described.

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 September 11, 2020.

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. Terminology	4
3. Upstream Selection Mechanism	5
3.1. Channel-Based Upstream Selection	5
3.2. Subscriber-Based Upstream Selection	5
3.3. Multiple Upstream Interface Selection for Robust Data Reception	5
4. Candidate Upstream Interface Configuration	6
4.1. Address Prefix Record	6
4.2. Channel/Session ID	7
4.3. Interface Priority	7
4.4. Default Upstream Interface	8
5. Upstream Interface Takeover	8
5.1. Proxy Behavior	8
5.2. Active Interval	9
6. Automatic Upstream Interface Configuration	9
6.1. Signaling-based Upstream Interface Configuration	10
6.2. Controller-based Upstream Interface Configuration	10
7. IANA Considerations	11
8. Security Considerations	11
9. Consideration for Updating YANG Model	11
10. References	11
10.1. Normative References	11
10.2. Informative References	12
Authors' Addresses	12

1. Introduction

The Internet Group Management Protocol (IGMP) [2][4] for IPv4 and the Multicast Listener Discovery Protocol (MLD) [3][4] for IPv6 are the standard protocols for hosts to initiate joining or leaving of multicast sessions. A proxy device performing IGMP/MLD-based forwarding (as known as IGMP/MLD proxy) [5] maintains multicast membership information by IGMP/MLD protocols on the downstream interfaces and sends IGMP/MLD membership report messages via the upstream interface to the upstream multicast routers when the membership information changes (e.g., by receiving solicited/unsolicited report messages). The proxy device forwards appropriate multicast packets received on its upstream interface to each downstream interface based on the downstream interface's subscriptions.

According to the specification of [5], an IGMP/MLD proxy has **a single** upstream interface and one or more downstream interfaces. The multicast forwarding tree must be manually configured by designating upstream and downstream interfaces on an IGMP/MLD proxy device, and the root of the tree is expected to be connected to a wider multicast infrastructure. An IGMP/MLD proxy device hence performs the router portion of the IGMP or MLD protocol on its downstream interfaces, and the host portion of IGMP/MLD on its upstream interface. The proxy device must not perform the router portion of IGMP/MLD on its upstream interface.

On the other hand, there is a scenario in which an IGMP/MLD proxy device enables multiple upstream interfaces and receives multicast packets through these interfaces. For example, a proxy device having more than one interface may want to access to different networks, such as a global network like the Internet and local-scope networks. Or, a proxy device having wired link (e.g., ethernet) and high-speed wireless link (e.g., WiMAX or LTE) may want to have the capability to connect to the Internet through both links. These proxy devices shall receive multicast packets from the different upstream interfaces and forward to the downstream interface(s). Several other scenarios and subsequent requirements for the support of multiple upstream interfaces on IGMP/MLD proxy are documented in [7].

This document describes the mechanism that enables an IGMP/MLD proxy device to receive multicast sessions/channels through the different upstream interfaces. The mechanism is configured with either "channel-based upstream selection" or "subscriber-based upstream selection", or both of them. By channel-based upstream selection, an IGMP/MLD proxy device selects one or multiple upstream interface(s) from the candidate upstream interfaces "per channel/session". By subscriber-based upstream selection, an IGMP/MLD proxy device selects one or multiple upstream interface(s) from the candidate upstream interfaces "per subscriber/receiver".

When a proxy device transmits an IGMP/MLD report message, it examines the source and multicast addresses in the IGMP/MLD records of the report message. It then transmits the appropriate IGMP/MLD report message(s) from the selected upstream interface(s). When a proxy device selects "one" upstream interface from the candidate upstream interfaces per session/channel, it enables "load balancing" per session/channel. When a proxy device selects "more than two" upstream interfaces from the candidate upstream interfaces per session/channel, it potentially receives duplicate (redundant) packets for the session/channel from the different upstream interfaces simultaneously and provides "robust data reception".

A mechanism for "upstream interface takeover" is also described in this document; when the selected upstream interface is going down or the state of the link attached to the upstream interface is inactive, one of the other active candidate upstream interfaces takes over the upstream interface (if configured). The potential timer values to switch from an inactive upstream interface to an active upstream interface from a list of candidate upstream interfaces are discussed in this document as well.

An "automatic upstream configuration" mechanism that selects an appropriate upstream interface(s) for sessions/channels based on the network and adjacent routers' conditions is also described in this document.

2. Terminology

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 [1].

In addition, the following terms are used in this document.

Selected upstream interface (or simply, upstream interface):
A proxy device's interface in the direction of the root of the multicast forwarding tree. A proxy device performs the host portion of IGMP/MLD on its upstream interfaces. An upstream interface is selected from a list of candidate upstream interfaces.

Default upstream interface:
A default upstream interface is the upstream interface for multicast sessions/channels for which a proxy device cannot choose other interfaces as the upstream interface. A default upstream interface is configured.

Active upstream interface:
An active upstream interface is the upstream interface that has been receiving packets for specific multicast sessions/channels during the pre-defined active interval.

Inactive upstream interface:
An inactive upstream interface is the interface that has not received packets for specific multicast sessions/channels during the pre-defined active interval.

Downstream interface:
Each of a proxy device's interfaces that is not in the direction of the root of the multicast forwarding tree. A proxy device performs the router portion of IGMP/MLD on its downstream interfaces.

Candidate upstream interface:

An interface that potentially becomes an upstream interface of the proxy device. A list of candidate upstream interfaces is configured with subscriber address prefixes, channel/session IDs, and priority values on an IGMP/MLD proxy device.

Channel/session ID:

Channel/session ID consists of source address prefix and multicast address prefix for which a candidate upstream interface supposes to be an upstream interface for specified multicast sessions/channels. Both or either source address prefix and/or multicast address prefix can be "null".

3. Upstream Selection Mechanism

3.1. Channel-Based Upstream Selection

An IGMP/MLD proxy device selects one or multiple upstream interface(s) from the candidate upstream interfaces "per channel/session" based on the "channel/session ID" configuration. This mechanism is called "channel-based upstream selection" whose configuration is explained in Section 4.1 and Section 4.2). enables an IGMP/MLD proxy device to use one or multiple upstream interface(s) from the candidate upstream interfaces "per channel/session" based on the "channel/session ID" configuration (as will be in Section 4.1 and Section 4.2).

3.2. Subscriber-Based Upstream Selection

An IGMP/MLD proxy device selects one or multiple upstream interface(s) from the candidate upstream interfaces "per subscriber/receiver". This is called "subscriber-based upstream selection". It enables a proxy device to use one or multiple upstream interface(s) per session/channel from the "candidate upstream interfaces" based on the "subscriber address prefix" configuration (as will be in Section 4.1).

3.3. Multiple Upstream Interface Selection for Robust Data Reception

When more than one candidate upstream interface is configured with the same source and multicast addresses for the "channel/session IDs", and "interface priority values" (as will be in Section 4.3) are identical, these candidate upstream interfaces act as the upstream interfaces for the sessions/channels and receive the packets simultaneously. This multiple upstream interface selection implements duplicate packet reception from redundant paths. It may improve data reception quality or robustness for a session/channel, as the same multicast data packets can come from different upstream

interfaces simultaneously. However, this robust data reception does not guarantee that the packets come from disjoint paths. It only configures that the adjacent upstream routers are different.

4. Candidate Upstream Interface Configuration

Candidate upstream interfaces are the interfaces from which an IGMP/MLD proxy device selects as an upstream interface. The upstream interface selection works with the configurations of "subscriber address prefix", "channel/session ID", and "interface priority value".

4.1. Address Prefix Record

An IGMP/MLD proxy device can configure the "subscriber address prefix" and "channel/session ID" for each candidate upstream interface. Channel/session ID consists of "source address prefix" and "multicast address prefix". Subscriber address prefix and source address prefix MUST be a valid unicast address prefix, and multicast address prefix MUST be a valid multicast address prefix. A proxy selects an upstream interface from its candidate upstream interfaces based on the configuration of the following address prefix record:

(subscriber address prefix, (channel/session ID))

where channel/session ID includes:

(source address prefix, multicast address prefix)

The default values of these address prefixes are "null". Null source address prefix represents the wildcard source address prefix, which indicates any host. Null multicast address prefix represents the wildcard multicast address prefix, which indicates the entire multicast address range (i.e., '224.0.0.0/4' for IPv4 or 'ff00::/8' for IPv6).

The candidate upstream interface having the configuration of subscriber address prefix is prioritized. If network operators want to assign a specific upstream interface for specific subscribers without depending on source and multicast address prefixes, both source and multicast addresses in the address prefix record is configured "null".

If network operators want to select specific upstream interface(s) without depending on subscriber address prefix, subscriber address prefix in the address prefix record is configured "null".

4.2. Channel/Session ID

Channel/session ID configuration consists of source and multicast address prefixes. Both/either source and/or multicast address may be configured "null". A candidate upstream interface having non-null source and multicast address configuration is prioritized for the upstream interface selection. For example, if a proxy device has two candidate upstream interfaces for the same multicast address prefix and one of them has non-null source address configuration, then that candidate upstream interface is selected for the source and multicast address pair. The other candidate upstream interface is selected for the configured multicast address prefix except the source address configured by the prior interface.

Source address prefix configuration takes priority over multicast address prefix configuration. For example, consider the case that an IGMP/MLD proxy device has a configuration with source address prefix S_p for the candidate upstream interface A and multicast address prefix G_p for the candidate upstream interface B. When it deals with an IGMP/MLD record whose source address, let's say S , is in the range of S_p , and whose multicast address, let's say G , is in the range of G_p , the proxy device selects the candidate upstream interface A, which supports the source address prefix, as the upstream interface, and transmits the (S,G) record via the interface A.

The same address prefix may be configured on different candidate upstream interfaces. When the same address prefix is configured on different candidate upstream interfaces, an upstream interface for that address prefix is selected based on each interface priority value (as will be in Section 4.3).

4.3. Interface Priority

An IGMP/MLD proxy device can configure the "interface priority" value for each candidate upstream interface. It is an integer value and is part of the configuration. The default value of the interface priority is the lowest value.

The interface priority value effects only when either of the following conditions is satisfied.

- o None of the candidate upstream interfaces configures both the subscriber address prefix and the channel/session ID.
- o More than one candidate upstream interface configures the same channel/session IDs but does not configure the subscriber address prefix.

In these conditions, the candidate upstream interface with the highest priority is chosen as the upstream interface. And as stated in Section 3.3, if the priority values for candidate upstream interfaces are also identical, all of these interfaces act as the upstream interfaces for the configured channel/session ID and may receive duplicate packets.

4.4. Default Upstream Interface

An IGMP/MLD proxy device SHOULD configure "a default upstream interface" for all incoming sessions/channels. A default upstream interface is selected as the upstream interface, when none of the candidate upstream interfaces configures subscriber address prefix, channel/session ID, or interface priority value, or with either of the following conditions.

- o None of the candidate upstream interfaces configures both the subscriber address prefix, the channel/session ID, and identical interface priority value.
- o More than one candidate upstream interface configures the same channel/session IDs and identical interface priority value, but does not configure the subscriber address prefix.

If a default upstream interface is not configured on an IGMP/MLD proxy device, the candidate upstream interface whose IPv4/v6 address is the highest of others is configured as the default upstream interface for the proxy device.

5. Upstream Interface Takeover

5.1. Proxy Behavior

If a selected upstream interface is going down or inactive, or an adjacent upstream router is not working, the upstream interface can be disabled and the other active upstream interface listed in the candidate upstream interfaces covering the same channel/session ID can act as a new upstream interface. It recursively examines the list of the candidate upstream interfaces (except the disabled interface) and decides a new upstream interface from them. If no active candidate upstream interfaces exist, the default upstream interface takes its role.

This function called "upstream interface takeover" is a default function for a proxy device that enables multiple upstream interface support. If a proxy device simultaneously uses more than two upstream interfaces per session/channel, and one or some of these upstream interface(s) is/are inactive, the proxy device acts either

of the following behaviors based on the configuration; (1) it only uses the active upstream interface(s) and does not add (i.e., does not complement) other upstream interfaces, (2) it uses the active upstream interface(s) and another candidate upstream interface whose priority is highest among the configured upstream interfaces, or (3) it uses the active upstream interface(s) and the default upstream interface.

The condition whether the upstream adjacent router is active or not can be decided by checking the link/interface condition on the proxy device or detected by monitoring IGMP/MLD Query or PIM [6] Hello message reception on the link. There are the cases that PIM is not running on the link or IGMP/MLD Query messages are not always transmitted by the upstream router (e.g., because of enabling the explicit tracking function [8]). Therefore, network operators MUST configure either; (1) the proxy device disables the upstream interface takeover, (2) the proxy device triggers upstream interface takeover by detecting no IGMP/MLD Query message within the active interval, or (3) the proxy device triggers upstream interface takeover by detecting no PIM Hello message within the active interval, for each candidate upstream interface.

Network operators may want to keep out of use for the inactive upstream interface(s). This causes, for example, when subscriber-based upstream selection is configured, according to their accounting policy (because the specific subscribers are planned to use the specific upstream interface and cannot receive packets from other upstream interfaces.) In that case, this upstream interface takeover must be disabled, and the proxy device keeps using that interface as the upstream interface for them (and waits for working the interface later again).

5.2. Active Interval

Active interval is a period, after which a proxy device recognizes that the selected upstream interface is inactive. Active interval for each candidate upstream interface SHOULD be configured. The active interval values are different in the situation whether the network operators want to trigger by either IGMP/MLD or PIM messages. The default active interval to detect an inactive upstream interface is around twice of IGMP/MLD General Query interval and PIM Hello interval. Further discussion [TBD].

6. Automatic Upstream Interface Configuration

6.1. Signaling-based Upstream Interface Configuration

There may be the ways for a proxy device to automatically configure the upstream interface for specific multicast channels/sessions. It works for the case in which there are no static configurations for a candidate upstream interface or operators decide. The algorithms are achieved by monitoring existing or newly proposed IGMP/MLD messages, but further discussions are TBD.

6.2. Controller-based Upstream Interface Configuration

A centralized controller can instruct the proxy what upstream interface is the appropriate one to use based on a certain multicast channel or on the user herself.

There are options for selecting the most appropriate upstream interface:

- o Association of membership requests from a specific user, identified by the source IP of the IGMP/MLD message, to a specific upstream interface, meaning that all the multicast traffic for that end user is received from a certain upstream interface.
- o Association of (S,G) to a specific upstream interface, meaning that an end user request for specific content delivered from a specific source should be received from a certain upstream interface.
- o Association of (*,G) to a specific upstream interface, meaning that a user request of given content, independently of the source of that content, should be received from a certain upstream interface.
- o Association of (S,*) to a specific upstream interface, meaning that all the requests from a certain user, independently of the group identifying the content, should be received from a certain upstream interface.

The list above does not show any precedence. Thus precedence should be defined to indicate priority in the selection of the criteria to apply, as a list of ordered actions.

The controller should also configure a default upstream interface for those subscription requests that do not match an explicitly configured behavior. In case of upstream interface failure, the default upstream interface could take over the failed upstream to provide redundancy.

To enable this manner of configuration, some control and management interface has to be supported by the proxy in order to receive configuration instructions from the controller.

The controller could interact with a number of proxies in the network. Being a centralized element, it could take coordinated decisions for managing all the multicast traffic in the network in a coordinated manner.

7. IANA Considerations

This document has no actions for IANA.

8. Security Considerations

This document neither provides new functions nor modifies the standard functions defined in [2][3][4]; hence there is no additional security consideration provided for these protocols themselves. On the other hand, it may be possible to encounter DoS attacks to make the function for upstream interface takeover stop if attackers illegally sends IGMP/MLD Query or PIM Hello messages on a LAN within a shorter period (i.e., before expiring the active interval for the upstream interface). To bypass such threats, it is recommended to capture the source addresses of IGMP/MLD Query or PIM Hello message senders and check whether the addresses correspond to the correct adjacent upstream routers. Consideration [TBD].

9. Consideration for Updating YANG Model

About the IGMP/MLD YANG model proposed in [9], there is a description of interfaces for IGMP (similar for MLD). When this document is officially approved, it is necessary to update the proposed YANG model to include all the information related to the upstream interfaces defined in this document, and consider the actions related to the selection of the upstream interfaces as mentioned in Section 6.

10. References

10.1. Normative References

- [1] Bradner, S., "Key words for use in RFCs to indicate requirement levels", RFC 2119, March 1997.
- [2] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

- [3] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.
- [4] Liu, H., Cao, W., and H. Asaeda, "Lightweight Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Version 2 (MLDv2) Protocols", RFC 5790, February 2010.
- [5] Fenner, B., He, H., Haberman, B., and H. Sandick, "Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD)-Based Multicast Forwarding ("IGMP/MLD Proxying")", RFC 4605, August 2006.
- [6] 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)", RFC 7761, March 2016.

10.2. Informative References

- [7] Contreras, LM., Bernardos, CJ., Asaeda, H., and N. Leymann, "Requirements for the extension of the IGMP/MLD proxy functionality to support multiple upstream interfaces", draft-ietf-pim-multiple-upstreams-reqs-08 (work-in-progress), November 2018.
- [8] Asaeda, H., "IGMP/MLD-Based Explicit Membership Tracking Function for Multicast Routers", draft-ietf-pim-explicit-tracking-13 (work-in-progress), November 2015.
- [9] Liu, X., Guo, F., Sivakumar, M., McAllister, P., and A. Peter, "A YANG Data Model for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD)", draft-ietf-pim-igmp-mld-yang-15 (work-in-progress), June 2019.

Authors' Addresses

Hitoshi Asaeda
National Institute of Information and Communications Technology
4-2-1 Nukui-Kitamachi
Koganei, Tokyo 184-8795
Japan

Email: asaeda@nict.go.jp

Luis M. Contreras
Telefonica
Ronda de la Comunicacion, s/n
Sur-3 building, 3rd floor
Madrid 28050
Spain

Email: luismiguel.contrerasmurillo@telefonica.com
URI: <http://lmcontreras.com/>

PIM Working Group
Internet-Draft
Intended status: Standards Track
Expires: 12 June 2022

G. Mirsky
Ericsson
J. Xiaoli
ZTE Corporation
9 December 2021

Fast Failover in Protocol Independent Multicast - Sparse Mode (PIM-SM)
Using Bidirectional Forwarding Detection (BFD) for Multipoint Networks
draft-ietf-pim-bfd-p2mp-use-case-10

Abstract

This document specifies how Bidirectional Forwarding Detection for multipoint networks can provide sub-second failover for routers that participate in Protocol Independent Multicast - Sparse Mode (PIM-SM). An extension to the PIM Hello message used to bootstrap a point-to-multipoint BFD session is also defined in this document.

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 12 June 2022.

Copyright Notice

Copyright (c) 2021 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
1.1. Conventions used in this document	3
1.1.1. Terminology	3
1.1.2. Requirements Language	3
2. BFD Discriminator PIM Hello Option	3
2.1. Using P2MP BFD in PIM Router Monitoring	4
2.2. P2MP BFD in PIM DR Load Balancing	5
2.3. Multipoint BFD Encapsulation	5
3. IANA Considerations	6
4. Security Considerations	6
5. Acknowledgments	6
6. References	6
6.1. Normative References	6
6.2. Informative References	7
Authors' Addresses	7

1. Introduction

Faster convergence in the control plane minimizes the periods of traffic blackholing, transient routing loops, and other situations that may negatively affect service data flow. Faster convergence in the control plane is beneficial to unicast and multicast routing protocols.

[RFC7761] is the current specification of the Protocol Independent Multicast - Sparse Mode (PIM-SM) for IPv4 and IPv6 networks. A conforming implementation of PIM-SM elects a Designated Router (DR) on each PIM-SM interface. When a group of PIM-SM nodes is connected to a shared media segment, e.g., Ethernet, the node elected as DR acts on behalf of directly connected hosts in the context of the PIM-SM protocol. Failure of the DR impacts the quality of the multicast services it provides to directly connected hosts because the default failure detection interval for PIM-SM routers is 105 seconds.

Bidirectional Forwarding Detection (BFD) [RFC5880] was originally defined to detect a failure of a point-to-point (p2p) path, single-hop [RFC5881] or multihop [RFC5883]. In some PIM-SM deployments, a

p2p BFD can be used to detect a failure and enable faster failover. [RFC8562] extends the BFD base specification [RFC5880] for multipoint and multicast networks, which matches the deployment scenarios for PIM-SM over a LAN segment. A BFD system in p2mp environment that transmits BFD Control messages using the BFD Demand mode [RFC5880] creates less BFD state than the Asynchronous mode. Point-to-multipoint (p2mp) BFD can enable faster detection of PIM-SM router failure compared to PIM-SM without BFD and thus minimize multicast service disruption. The monitored PIM-SM router acts as the head and other routers as tails of a p2mp BFD session. This document defines the monitoring of a PIM-SM router using p2mp BFD. This document also defines the extension to PIM-SM [RFC7761] to bootstrap a PIM-SM router to join in p2mp BFD session over shared media segment.

1.1. Conventions used in this document

1.1.1. Terminology

This document uses terminology defined in [RFC5880], [RFC8562], and [RFC7761]. Familiarity with these specifications and the terminology used is expected.

1.1.2. 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.

2. BFD Discriminator PIM Hello Option

Figure 1 displays the new optional BFD Discriminator PIM Hello option to bootstrap a tail of the p2mp BFD session.

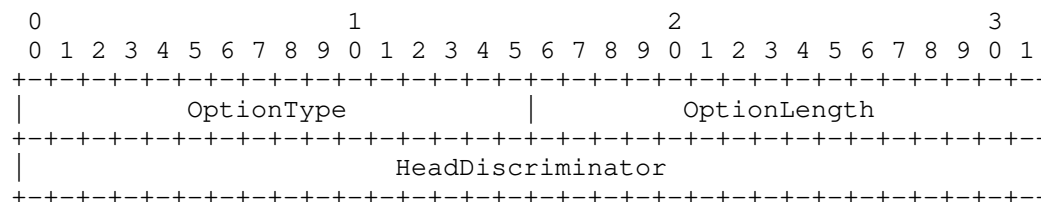


Figure 1: BFD Discriminator PIM Hello Option

where new fields are interpreted as:

OptionType: TBA.

OptionLength: MUST be set to 4.

HeadDiscriminator: the four-octet field MUST be included in the BFD Discriminator PIM-SM Hello option. The value MUST NOT be zero. It equals the value of My Discriminator ([RFC5880]) allocated by the head.

If the value of the OptionLength field is not equal to 4, the BFD Discriminator PIM Hello option is considered malformed, and the receiver MUST stop processing PIM Hello options. If the value of the HeadDiscriminator field equals zero, then the BFD Discriminator PIM Hello option MUST be considered invalid, and the receiver MUST ignore it. The receiver SHOULD log a notification regarding the malformed or invalid BFD Discriminator Hello option under the control of a throttling logging mechanism.

2.1. Using P2MP BFD in PIM Router Monitoring

If the head is no longer serving the function that prompted it to be monitored, then it MUST cease including the BFD Discriminator PIM Hello option in its PIM-Hello message, and it SHOULD shut down the BFD session following the procedures described in Section 5.9 [RFC8562].

The head MUST create a BFD session of type MultipointHead [RFC8562]. Note that any PIM-SM router, regardless of its role, MAY become a head of a p2mp BFD session. To control the volume of BFD control traffic on a shared media segment, an operator should carefully select PIM-SM routers configured as a head of a p2mp BFD session. The head MUST include the BFD Discriminator PIM Hello option in its PIM Hello messages.

A PIM-SM router that is configured to monitor the head by using p2mp BFD is referred to throughout this document as a "tail". When such a tail receives a PIM-Hello packet with the BFD Discriminator PIM Hello option, the tail MAY create a p2mp BFD session of type MultipointTail, as defined in [RFC8562].

The node that includes the BFD Discriminator PIM Hello option transmits BFD Control packets periodically. For the tail to correctly demultiplex BFD [RFC8562], the source address and My Discriminator of the BFD packets MUST be the same as the source address and the HeadDiscriminator, respectively, of the PIM Hello message. If that is not the case, the tail BFD node would not be able to monitor the state of the PIM-SM node, that is, the head of the p2mp BFD session, though the regular PIM-SM mechanisms remain fully operational.

If the tail detects a MultipointHead failure [RFC8562], it MUST delete the corresponding neighbor state and follow procedures defined in [RFC7761] for the DR and additional neighbor state deletion after the neighbor timeout expires.

If the head ceases to include the BFD Discriminator PIM Hello option in its PIM-Hello message, tails SHOULD close the corresponding MultipointTail BFD session without affecting the PIM state in any way. Thus, the tail stops using BFD to monitor the head and reverts to the procedures defined in [RFC7761].

2.2. P2MP BFD in PIM DR Load Balancing

[RFC8775] specifies the PIM Designated Router Load Balancing (DRLB) functionality. Any PIM router that advertises the DRLB-Cap Hello Option can become the head of a p2mp BFD session, as specified in Section 2.1. The head router administratively sets the bfd.SessionState to Up in the MultipointHead session [RFC8562] only if it is a Group Designated Router (GDR) Candidate, as specified in Sections 5.5 and 5.6 of [RFC8775]. If the router is no longer the GDR, then it MUST shut down following the procedures described in Section 5.9 [RFC8562]. For each GDR Candidate that includes BFD Discriminator option in its PIM Hello, the PIM DR MUST create a MultipointTail session [RFC8562]. PIM DR demultiplexes BFD sessions based on the value of the My Discriminator field and the source IP address. If PIM DR detects a failure of one of the sessions, it MUST remove that router from the GDR Candidate list and immediately transmit a new DRLB-List option.

2.3. Multipoint BFD Encapsulation

The MultipointHead of a p2mp BFD session when transmitting BFD Control packets:

MUST set TTL or Hop Limit value to 255 (Section 5 [RFC5881]). Similarly, all received BFD Control packets that are demultiplexed to the session MUST be discarded if the received TTL or Hop Limit is not equal to 255;

MUST use the group address ALL-PIM-ROUTERS ('224.0.0.13' for IPv4 and 'ff02::d' for IPv6) as destination IP address

3. IANA Considerations

IANA is requested to allocate a new OptionType value from PIM-Hello Options registry according to:

Value	Length	Name	Reference
TBA	4	BFD Discriminator Option	This document

Table 1: BFD Discriminator option type

4. Security Considerations

This document defines a way to accelerate detecting a failure that affects PIM functionality by using BFD. The operation of either protocol is not changed.

The security considerations discussed in [RFC7761], [RFC5880], [RFC5881], [RFC8562], and [RFC8775] apply to this document.

5. Acknowledgments

The authors cannot say enough to express their appreciation of the comments and suggestions we received from Stig Venaas. The authors greatly appreciate the comments and suggestions by Alvaro Retana that improved the clarity of this document.

6. References

6.1. Normative References

- [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>>.
- [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>>.

- [RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, DOI 10.17487/RFC5881, June 2010, <<https://www.rfc-editor.org/info/rfc5881>>.
- [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>>.
- [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>>.
- [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>>.
- [RFC8775] Cai, Y., Ou, H., Vallepalli, S., Mishra, M., Venaas, S., and A. Green, "PIM Designated Router Load Balancing", RFC 8775, DOI 10.17487/RFC8775, April 2020, <<https://www.rfc-editor.org/info/rfc8775>>.

6.2. Informative References

- [RFC5883] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for Multihop Paths", RFC 5883, DOI 10.17487/RFC5883, June 2010, <<https://www.rfc-editor.org/info/rfc5883>>.

Authors' Addresses

Greg Mirsky
Ericsson

Email: gregimirsky@gmail.com

Ji Xiaoli
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing,
China

Email: ji.xiaoli@zte.com.cn

PIM Working Group
Internet Draft
Intended status: Standards Track
Expires: February 28, 2022

H. Zhao
Ericsson
X. Liu
Volta
Y. Liu
China Mobile
M. Panchanathan
Cisco
M. Sivakumar
Juniper

August 30, 2021

A Yang Data Model for IGMP/MLD Proxy
draft-ietf-pim-igmp-mld-proxy-yang-06.txt

Abstract

This document defines a YANG data model that can be used to configure and manage Internet Group Management Protocol (IGMP) or Multicast Listener Discovery (MLD) proxy devices. The YANG module in this document conforms to Network Management Datastore Architecture (NMDA).

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on February 28, 2022.

Copyright Notice

Copyright (c) 2021 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 (<http://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.....	3
1.1. Terminology.....	3
1.2. Conventions Used in This Document.....	3
1.3. Tree Diagrams.....	4
1.4. Prefixes in Data Node Names.....	4
2. Design of Data Model.....	4
2.1. Overview.....	5
2.2. Optional Capabilities.....	5
2.3. Position of Address Family in Hierarchy.....	5
3. Module Structure.....	6
3.1. IGMP Proxy Configuration and Operational State.....	6
3.2. MLD Proxy Configuration and Operational State.....	7
4. IGMP/MLD Proxy YANG Module.....	8
5. Security Considerations.....	15
6. IANA Considerations.....	16
6.1. XML Registry.....	16
6.2. YANG Module Names Registry.....	16
7. References.....	17
7.1. Normative References.....	17
7.2. Informative References.....	18
Appendix. Data Tree Example.....	19
Authors' Addresses.....	22

1. Introduction

This document defines a YANG [RFC7950] data model for the management of Internet Group Management Protocol (IGMP) or Multicast Listener Discovery (MLD) Proxy [RFC4605] devices.

The YANG module in this document conforms to the Network Management Datastore Architecture defined in [RFC8342]. The "Network Management Datastore Architecture" (NMDA) adds the ability to inspect the current operational values for configuration, allowing clients to use identical paths for retrieving the configured values and the operational values.

1.1. Terminology

The terminology for describing YANG data models is found in [RFC6020] and [RFC7950], including:

- * augment
- * data model
- * data node
- * identity
- * module

The following abbreviations are used in this document and defined model:

IGMP: Internet Group Management Protocol [RFC3376].

MLD: Multicast Listener Discovery [RFC3810].

PIM: Protocol Independent Multicast [RFC7761].

1.2. Conventions Used in This Document

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.

1.3. Tree Diagrams

Tree diagrams used in this document follow the notation defined in [RFC8340].

1.4. Prefixes in Data Node Names

In this document, names of data nodes, and other data model objects are often used without a prefix, as long as it is clear from the context in which YANG module each name is defined. Otherwise, names are prefixed using the standard prefix associated with the corresponding YANG module, as shown in Table 1.

Prefix	YANG module	Reference
inet	ietf-inet-types	[RFC6991]
if	ietf-interfaces	[RFC8343]
rt	ietf-routing	[RFC8349]
rt-types	ietf-routing-types	[RFC8294]
pim-base	ietf-pim-base	[draft-ietf-pim-yang]

Table 1: Prefixes and Corresponding YANG Modules

2. Design of Data Model

The model covers Considerations for Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD) - Based Multicast Forwarding ("IGMP/MLD Proxying") [RFC4605].

The goal of this document is to define a data model that provides a common user interface to IGMP/MLD Proxy. This document provides freedom for vendors to adapt the data model to their product implementations.

2.1. Overview

The model defined in this document has all the common building blocks for the IGMP/MLD Proxy devices. It can be used to configure IGMP/MLD Proxy. The operational state data and statistics can also be retrieved by it.

The model contains all the basic configuration parameters. The occasionally implemented parameters are modeled as optional features in this model, while the rarely implemented parameters are not included in this model and left for augmentation.

2.2. Optional Capabilities

This model is designed to represent the basic capability subsets of IGMP / MLD Proxy. The main design goals of this document are that the basic capabilities described in the model are supported by any major now-existing implementation, and that the configuration of all implementations meeting the specifications is easy to express through some combination of the optional features in the model and simple vendor augmentations.

There is also value in widely supported features being standardized, to provide a standardized way to access these features, to save work for individual vendors, and so that mapping between different vendors' configuration is not needlessly complicated. Therefore, this model declares a number of features representing capabilities that not all deployed devices support.

The extensive use of feature declarations should also substantially simplify the capability negotiation process for a vendor's IGMP / MLD Proxy implementations.

2.3. Position of Address Family in Hierarchy

IGMP Proxy only supports IPv4, while MLD Proxy only supports IPv6. The data model defined in this document can be used for both IPv4 and IPv6 address families.

This document defines IGMP Proxy and MLD Proxy as separate schema branches in the structure. The benefits are:

- * The model can support IGMP Proxy (IPv4), MLD Proxy (IPv6), or both optionally and independently. Such flexibility cannot be achieved cleanly with a combined branch.

* The structure is consistent with other YANG data models such as [RFC8652], which uses separate branches for IPv4 and IPv6.

* Having separate branches for IGMP Proxy and MLD Proxy allows minor differences in their behavior to be modelled more simply and cleanly. The two branches can better support different features and node types.

3. Module Structure

This model augments the core routing data model specified in [RFC8349].

```
+--rw routing
  +--rw router-id?
  +--rw control-plane-protocols
    |   +--rw control-plane-protocol* [type name]
    |   |   +--rw type
    |   |   +--rw name
    |   |   +--rw igmp-proxy <= Augmented by this Model
    |   |   ...
    |   +--rw mld-proxy <= Augmented by this Model
```

The "igmp-proxy" container instantiates IGMP Proxy. The "mld-proxy" container instantiates MLD Proxy.

The YANG data model defined in this document conforms to the Network Management Datastore Architecture (NMDA) [RFC8342]. The operational state data is combined with the associated configuration data in the same hierarchy [RFC8407].

3.1. IGMP Proxy Configuration and Operational State

The YANG module augments /rt:routing/rt:control-plane-protocols/rt:control-plane-protocol to add the igmp-proxy container.

All the IGMP Proxy related attributes are defined in the igmp-proxy container. The read-write attributes represent configurable data. The read-only attributes represent state data.

The igmp-version represents version of IGMP protocol, and default value is 2. If the value of enable is true, it means IGMP Proxy is enabled.

The interface list under igmp-proxy contains upstream interfaces for IGMP proxy. There is also a constraint to make sure the upstream interface for IGMP proxy should not be configured PIM.

To configure a downstream interface for IGMP proxy, it is needed to enable IGMP on that interface. This is defined in the YANG Data Model

for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) [RFC8652].

```

augment /rt:routing/rt:control-plane-protocols
  /rt:control-plane-protocol:
    +--rw igmp-proxy {igmp-proxy}?
      +--rw interfaces
        +--rw interface* [interface-name]
          +--rw interface-name          if:interface-ref
          +--rw igmp-version?           uint8
          +--rw enable?                 boolean
          +--rw sender-source-address?  inet:ipv4-address
          +--ro group* [group-address]
            +--ro group-address
              | rt-types:ipv4-multicast-group-address
            +--ro up-time?               uint32
            +--ro filter-mode            enumeration
            +--ro source* [source-address]
              +--ro source-address       inet:ipv4-address
              +--ro up-time?             uint32
              +--ro downstream-interface* [interface-name]
                +--ro interface-name    if:interface-ref

```

3.2. MLD Proxy Configuration and Operational State

The YANG module augments /rt:routing/rt:control-plane-protocols/rt:control-plane-protocol to add the mld-proxy container.

All the MLD Proxy related attributes are defined in the mld-proxy container. The read-write attributes represent configurable data. The read-only attributes represent state data.

The mld-version represents version of MLD protocol, and default value is 2. If the value of enable is true, it means MLD Proxy is enabled.

The interface list under mld-proxy contains upstream interfaces for MLD proxy. There is also a constraint to make sure the upstream interface for MLD proxy should not be configured PIM.

To configure a downstream interface for MLD proxy, enable MLD on that interface. This is defined in the YANG Data Model for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) [RFC8652].

```

augment /rt:routing/rt:control-plane-protocols
  /rt:control-plane-protocol:
    +--rw mld-proxy {mld-proxy}?
      +--rw interfaces
        +--rw interface* [interface-name]

```

```

+--rw interface-name          if:interface-ref
+--rw mld-version?            uint8
+--rw enable?                 boolean
+--rw sender-source-address?  inet:ipv6-address
+--ro group* [group-address]
  +--ro group-address
    |   rt-types:ipv6-multicast-group-address
  +--ro up-time?              uint32
  +--ro filter-mode           enumeration
  +--ro source* [source-address]
    +--ro source-address      inet:ipv6-address
    +--ro up-time?            uint32
    +--ro downstream-interface* [interface-name]
      +--ro interface-name    if:interface-ref

```

4. IGMP/MLD Proxy YANG Module

This module references [RFC4605], [RFC6991], [RFC8294], [RFC8343], [RFC8349] and [draft-ietf-pim-yang].

```

<CODE BEGINS> file ietf-igmp-mld-proxy@2021-04-21.yang
module ietf-igmp-mld-proxy {
  yang-version 1.1;
  namespace "urn:ietf:params:xml:ns:yang:ietf-igmp-mld-proxy";
  // replace with IANA namespace when assigned
  prefix igmp-mld-proxy;

  import ietf-inet-types {
    prefix inet;
  }
  import ietf-interfaces {
    prefix if;
  }
  import ietf-routing {
    prefix rt;
  }
  import ietf-routing-types {
    prefix rt-types;
  }
  import ietf-pim-base {
    prefix pim-base;
  }

  organization
    "IETF PIM Working Group";

  contact
    "WG Web:  <http://tools.ietf.org/wg/pim/>
    WG List:  <mailto:pim@ietf.org>

```

Editors: Hongji Zhao
<mailto:hongji.zhao@ericsson.com>

Xufeng Liu
<mailto:xufeng.liu.ietf@gmail.com>

Yisong Liu
<mailto:liuyisong@chinamobile.com>

Mani Panchanathan
<mailto:mapancha@cisco.com>

Mahesh Sivakumar
<mailto:sivakumar.mahesh@gmail.com>

";

description

"The module defines a collection of YANG definitions common for all Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxy devices.

Copyright (c) 2021 IETF Trust and the persons identified as authors of the code. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, is permitted pursuant to, and subject to the license terms contained in, the Simplified BSD License set forth in Section 4.c of the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>).

This version of this YANG module is part of RFC XXXX; see the RFC itself for full legal notices.";

```
revision 2021-04-21 {  
  description  
    "Initial revision.";  
  reference  
    "RFC XXXX: A YANG Data Model for IGMP and MLD Proxy";  
}
```

```
/*  
 * Features  
 */
```

```
feature igmp-proxy {  
  description  
    "Support IGMP Proxy protocol.";  
  reference  
    "RFC 4605";
```

```
}

feature mld-proxy {
  description
    "Support MLD Proxy protocol.";
  reference
    "RFC 4605";
}

/*
 * Identities
 */

identity igmp-proxy {
  base rt:control-plane-protocol;
  description
    "IGMP Proxy protocol";
}

identity mld-proxy {
  base rt:control-plane-protocol;
  description
    "MLD Proxy protocol";
}

/*
 * Groupings
 */

grouping per-interface-config-attributes {
  description "Config attributes under interface view";
  leaf enable {
    type boolean;
    default false;
    description
      "Set the value to true to enable IGMP/MLD proxy";
  }
} // per-interface-config-attributes

grouping state-group-attributes {
  description
    "State group attributes";
  leaf up-time {
    type uint32;
    units seconds;
    description
      "The elapsed time for (S,G) or (*,G).";
  }
  leaf filter-mode {
    type enumeration {
      enum "include" {
```

```
        description
            "In include mode, reception of packets sent
            to the specified multicast address is requested
            only from those IP source addresses listed in the
            source-list parameter";
    }
    enum "exclude" {
        description
            "In exclude mode, reception of packets sent
            to the given multicast address is requested
            from all IP source addresses except those
            listed in the source-list parameter.";
    }
}
mandatory true;
description
    "Filter mode for a multicast group,
    may be either include or exclude.";
}
} // state-group-attributes

/* augments */

augment "/rt:routing/rt:control-plane-protocols"+
    "/rt:control-plane-protocol" {
    when
        "derived-from-or-self(rt:type, 'igmp-mld-proxy:igmp-proxy')" {
            description
                "This augmentation is only valid for IGMP Proxy.";
        }
    description
        "IGMP Proxy augmentation to routing control plane protocol
        configuration and state.";
    container igmp-proxy {
        if-feature "igmp-proxy";
        description "IGMP proxy";
        container interfaces {
            description
                "Containing a list of upstream interfaces.";
            list interface {
                key "interface-name";
                description
                    "List of upstream interfaces.";
                leaf interface-name {
                    type if:interface-ref;
                    must "not( current() = /rt:routing"+
                        "/rt:control-plane-protocols/pim-base:pim"+
                        "/pim-base:interfaces/pim-base:interface"+
                        "/pim-base:name )" {
                        description
```

```
        "The upstream interface for IGMP proxy
        should not be configured PIM.";
    }
    description "The upstream interface name.";
}
leaf igmp-version {
    type uint8 {
        range "1..3";
    }
    default 2;
    description "IGMP version.";
}
uses per-interface-config-attributes;
leaf sender-source-address {
    type inet:ipv4-address;
    description
        "The sender source address of
        IGMP membership report or leave.";
}
list group {
    key "group-address";
    config false;
    description
        "Multicast group membership information
        that joined on the interface.";
    leaf group-address {
        type rt-types:ipv4-multicast-group-address;
        description
            "Multicast group address.";
    }
    uses state-group-attributes;
    list source {
        key "source-address";
        description
            "List of multicast source information
            of the multicast group.";
        leaf source-address {
            type inet:ipv4-address;
            description
                "Multicast source address";
        }
    }
    leaf up-time {
        type uint32;
        units seconds;
        description
            "The elapsed time for (S,G) or (*,G).";
    }
    list downstream-interface {
        key "interface-name";
        description "The downstream interfaces list.";
        leaf interface-name {
```



```
        type if:interface-ref;
        description
            "Downstream interfaces
             for each upstream-interface";
    }
    }
    } // list source
    } // list group
    } // interface
    } // interfaces
}
}

augment "/rt:routing/rt:control-plane-protocols"+
    "/rt:control-plane-protocol" {
    when
        "derived-from-or-self(rt:type, 'igmp-mld-proxy:mld-proxy')" {
        description
            "This augmentation is only valid for MLD Proxy.";
        }
    description
        "MLD Proxy augmentation to routing control plane protocol
         configuration and state.";
    container mld-proxy {
        if-feature "mld-proxy";
        description "MLD proxy";
        container interfaces {
            description
                "Containing a list of upstream interfaces.";
            list interface {
                key "interface-name";
                description
                    "List of upstream interfaces.";
                leaf interface-name {
                    type if:interface-ref;
                    must "not( current() = /rt:routing"+
                        "/rt:control-plane-protocols/pim-base:pim"+
                        "/pim-base:interfaces/pim-base:interface"+
                        "/pim-base:name )" {
                        description
                            "The upstream interface for MLD proxy
                             should not be configured PIM.";
                    }
                }
                description "The upstream interface name.";
            }
            leaf mld-version {
                type uint8 {
                    range "1..2";
                }
                default 2;
                description "MLD version.";
            }
        }
    }
}
```

```
    }
    uses per-interface-config-attributes;
    leaf sender-source-address {
        type inet:ipv6-address;
        description
            "The sender source address of
             MLD membership report or leave.";
    }
    list group {
        key "group-address";
        config false;
        description
            "Multicast group membership information
             that joined on the interface.";
        leaf group-address {
            type rt-types:ipv6-multicast-group-address;
            description
                "Multicast group address.";
        }
    }
    uses state-group-attributes;
    list source {
        key "source-address";
        description
            "List of multicast source information
             of the multicast group.";
        leaf source-address {
            type inet:ipv6-address;
            description
                "Multicast source address";
        }
        leaf up-time {
            type uint32;
            units seconds;
            description
                "The elapsed time for (S,G) or (*,G).";
        }
        list downstream-interface {
            key "interface-name";
            description "The downstream interfaces list.";
            leaf interface-name {
                type if:interface-ref;
                description
                    "Downstream interfaces
                     for each upstream-interface";
            }
        }
    } // list source
} // list group
} // interface
} // interfaces
}
```

```
    }  
  }  
<CODE ENDS>
```

5. Security Considerations

The YANG module specified in this document defines a schema for data that is designed to be accessed via network management protocols such as NETCONF [RFC6241] or RESTCONF [RFC8040]. The lowest NETCONF layer is the secure transport layer, and the mandatory-to-implement secure transport is Secure Shell (SSH) [RFC6242]. The lowest RESTCONF layer is HTTPS, and the mandatory-to-implement secure transport is TLS [RFC8446].

The Network Configuration Access Control Model (NACM) [RFC8341] provides the means to restrict access for particular NETCONF or RESTCONF users to a preconfigured subset of all available NETCONF or RESTCONF protocol operations and content.

There are a number of data nodes defined in this YANG module that are writable/creatable/deletable (i.e., config true, which is the default). These data nodes may be considered sensitive or vulnerable in some network environments. Write operations (e.g., edit-config) to these data nodes without proper protection can have a negative effect on network operations. These are the subtrees and data nodes and their sensitivity/vulnerability:

Under /rt:routing/rt:control-plane-protocols/rt:control-plane-protocol:/

igmp-mld-proxy:igmp-proxy

igmp-mld-proxy:mld-proxy

Unauthorized access to any data node of these subtrees can adversely affect the IGMP / MLD Proxy subsystem of both the local device and the network. This may lead to network malfunctions, delivery of packets to inappropriate destinations, and other problems.

Some of the readable data nodes in this YANG module may be considered sensitive or vulnerable in some network environments. It is thus important to control read access (e.g., via get, get-config, or notification) to these data nodes. These are the subtrees and data nodes and their sensitivity/vulnerability:

Under /rt:routing/rt:control-plane-protocols/rt:control-plane-protocol:/

igmp-mld-proxy:igmp-proxy

igmp-mld-proxy:mld-proxy

Unauthorized access to any data node of these subtrees can disclose the operational state information of IGMP / MLD Proxy on this device. The group/source information may expose multicast group memberships.

6. IANA Considerations

RFC Ed.: In this section, replace all occurrences of 'XXXX' with the actual RFC number (and remove this note).

6.1. XML Registry

This document registers the following namespace URIs in the IETF XML registry [RFC3688]:

URI: urn:ietf:params:xml:ns:yang:ietf-igmp-mld-proxy
Registrant Contact: The IETF.
XML: N/A, the requested URI is an XML namespace.

6.2. YANG Module Names Registry

This document registers the following YANG modules in the YANG Module Names registry [RFC7950]:

name:	ietf-igmp-mld-proxy
namespace:	urn:ietf:params:xml:ns:yang:ietf-igmp-mld-proxy
prefix:	igmp-mld-proxy
reference:	RFC XXXX

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.
- [RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.
- [RFC3688] Mealling, M., "The IETF XML Registry", RFC 3688, January 2004.
- [RFC3810] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.
- [RFC4605] B. Fenner, H. He, B. Haberman and H. Sandick, "Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD) - Based Multicast Forwarding ("IGMP/MLD Proxying")", RFC 4605, August 2006.
- [RFC6020] Bjorklund, M., Ed., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)", RFC 6020, October 2010.
- [RFC6241] R. Enns, Ed., M. Bjorklund, Ed., J. Schoenwaelder, Ed., A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, June 2011.
- [RFC6242] Wasserman, M., "Using the NETCONF Protocol over Secure Shell (SSH)", RFC 6242, June 2011.
- [RFC6991] Schoenwaelder, J., Ed., "Common YANG Data Types", RFC 6991, July 2013.
- [RFC7950] M. Bjorklund, Ed., "The YANG 1.1 Data Modeling Language", RFC 7950, August 2016.
- [RFC8040] A. Bierman, M. Bjorklund, K. Watsen, "RESTCONF Protocol", RFC 8040, January 2017.
- [RFC8174] B. Leiba, "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", RFC 8174, May 2017.
- [RFC8294] X. Liu, Y. Qu, A. Lindem, C. Hopps, L. Berger, "Common YANG Data Types for the Routing Area", RFC 8294, December 2017.
- [RFC8340] M. Bjorklund, and L. Berger, Ed., "YANG Tree Diagrams", RFC 8340, March 2018.

- [RFC8341] Bierman, A. and M. Bjorklund, "Network Configuration Access Control Model", RFC 8341, March 2018.
- [RFC8342] M. Bjorklund and J. Schoenwaelder, "Network Management Datastore Architecture (NMDA)", RFC 8342, March 2018.
- [RFC8343] M. Bjorklund, "A YANG Data Model for Interface Management", RFC 8343, March 2018.
- [RFC8349] L. Lhotka, A. Lindem, Y. Qu, "A YANG Data Model for Routing Management (NMDA Version)", RFC 8349, March 2018.
- [RFC8407] A. Bierman, "Guidelines for Authors and Reviewers of Documents Containing YANG Data Models", RFC 8407, October 2018.
- [RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, August 2018.
- [RFC8652] X. Liu, F. Guo, M. Sivakumar, P. McAllister, A. Peter, "A YANG Data Model for the Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD)", RFC 8652, November 2019.
- [draft-ietf-pim-yang] X. Liu, P. McAllister, A. Peter, M. Sivakumar, Y. Liu, F. Hu, "A YANG Data Model for Protocol Independent Multicast (PIM)", draft-ietf-pim-yang-17 (RFC Editor state is MISSREF), May 2018.

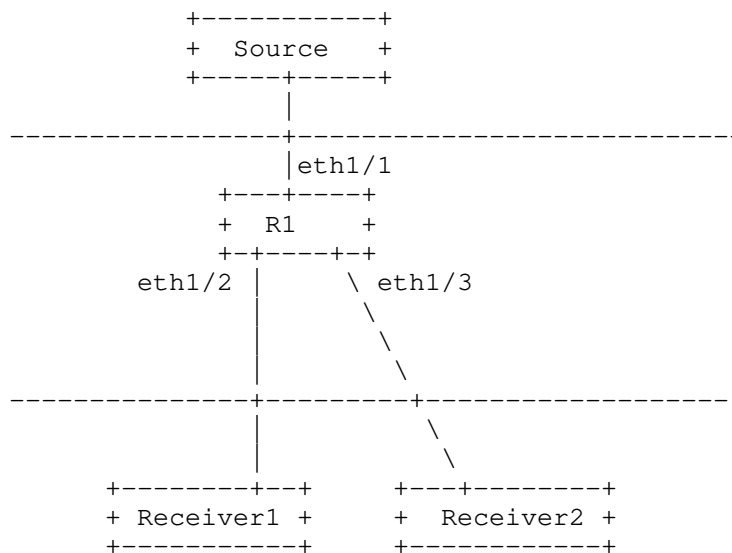
7.2. Informative References

- [RFC7761] B. Fenner, M. Handley, H. Holbrook, I. Kouvelas, R. Parekh, Z. Zhang, L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 7761, March 2016.
- [RFC7951] L. Lhotka, "JSON Encoding of Data Modeled with YANG", RFC 7951, August 2016.

Appendix. Data Tree Example

This section contains an example for IGMP Proxy in the JSON encoding [RFC7951], containing both configuration and state data. In the example IGMP Proxy is enabled on interface eth1/1.

It is also needed to enable IGMP on eth1/2 and eth1/3. The configuration details are omitted here because this document is focused on IGMP/MLD Proxy.



The configuration data for R1 in the above figure could be as follows:

```

{
  "ietf-interfaces:interfaces": {
    "interface": [
      {
        "name": "eth1/1",
        "type": "iana-if-type:ipForward",
        "ietf-ip:ipv4": {
          "address": [
            {
              "ip": "11.0.0.1",
              "prefix-length": 24
            }
          ]
        }
      }
    ]
  }
}

```

```

    ]
  },
  "ietf-routing:routing": {
    "control-plane-protocols": {
      "control-plane-protocol": [
        {
          "type": "ietf-igmp-mld-proxy:igmp-proxy",
          "name": "proxy1",
          "ietf-igmp-mld-proxy:igmp-proxy": {
            "interfaces": {
              "interface": [
                {
                  "interface-name": "eth1/1",
                  "igmp-version": 3,
                  "enable": true
                }
              ]
            }
          }
        }
      ]
    }
  }
}

```

The corresponding operational state data for R1 could be as follows:

```

{
  "ietf-interfaces:interfaces": {
    "interface": [
      {
        "name": "eth1/1",
        "type": "iana-if-type:ipForward",
        "admin-status": "up",
        "oper-status": "up",
        "if-index": 25678136,
        "statistics": {
          "discontinuity-time": "2021-05-23T10:34:56-06:00"
        },
        "ietf-ip:ipv4": {
          "address": [
            {
              "ip": "11.0.0.1",
              "prefix-length": 24
            }
          ]
        }
      }
    ]
  }
}

```



```

"ietf-routing:routing": {
  "control-plane-protocols": {
    "control-plane-protocol": [
      {
        "type": "ietf-igmp-mld-proxy:igmp-proxy",
        "name": "proxy1",
        "ietf-igmp-mld-proxy:igmp-proxy": {
          "interfaces": {
            "interface": [
              {
                "interface-name": "eth1/1",
                "igmp-version": 3,
                "enable": true,
                "group": [
                  {
                    "group-address": "225.0.0.1",
                    "filter-mode": "include",
                    "source": [
                      {
                        "source-address": "1.1.1.1",
                        "downstream-interface": [
                          {
                            "interface-name": "eth1/2"
                          },
                          {
                            "interface-name": "eth1/3"
                          }
                        ]
                      }
                    ]
                  }
                ]
              }
            ]
          }
        }
      }
    ]
  }
}

```

Authors' Addresses

Hongji Zhao
Ericsson (China) Communications Company Ltd.
Ericsson Tower, No. 5 Lize East Street,
Chaoyang District Beijing 100102, China
Email: hongji.zhao@ericsson.com

Xufeng Liu
Volta Networks
USA
EMail: Xufeng.liu.ietf@gmail.com

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

Mani Panchanathan
Cisco
India
Email: mapancha@cisco.com

Mahesh Sivakumar
Juniper Networks
1133 Innovation Way
Sunnyvale, California
USA
EMail: sivakumar.mahesh@gmail.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 11 May 2022

V. Kamath
VMware
R. Chokkanathapuram Sundaram
Cisco Systems, Inc.
R. Banthia
Apstra
A. Gopal
Cisco Systems, Inc.
7 November 2021

PIM Null-Register packing
draft-ietf-pim-null-register-packing-11

Abstract

In PIM-SM networks PIM Null-Register messages are sent by the Designated Router (DR) to the Rendezvous Point (RP) to signal the presence of Multicast sources in the network. There are periodic PIM Null-Registers sent from the DR to the RP to keep the state alive at the RP as long as the source is active. The PIM Null-Register message carries information about a single Multicast source and group.

This document defines a standard to send multiple Multicast source and group information in a single PIM Packed Null-Register message. We will refer to the new packed formats as the PIM Packed Null-Register format and PIM Packed Register-Stop format throughout the document. This document also discusses interoperability between the PIM routers which do not understand the PIM Packed Null-Register format and routers which do understand it.

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 11 May 2022.

Copyright Notice

Copyright (c) 2021 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
1.1. Conventions used in this document	3
1.2. Terminology	3
2. Packed Null-Register Capability	3
3. PIM Packed Null-Register message format	4
4. PIM Packed Register-Stop message format	5
5. Protocol operation	6
6. Operational Considerations	7
7. PIM Anycast RP Considerations	7
8. PIM RP router version downgrade	7
9. Fragmentation Considerations	7
10. Security Considerations	8
11. IANA Considerations	8
12. Acknowledgments	8
13. References	8
13.1. Normative References	8
13.2. Informative References	9
Authors' Addresses	9

1. Introduction

PIM Null-Registers are sent by the DR periodically for Multicast streams to keep the states active on the RP, as long as the multicast source is alive. As the number of multicast sources increases, the number of PIM Null-Register messages that are sent also increases. This results in more PIM packet processing at the RP and the DR.

The control plane policing (COPP), monitors the packets that are processed by the control plane. The high rate at which Null-Registers are received at the RP can lead to COPP drops of Multicast PIM Null-Register messages. This draft proposes a method to efficiently pack multiple PIM Null-Registers [RFC7761] (Section 4.4) and Register-Stops [RFC7761] (Section 3.2) into a single message as these packets anyway do not contain encapsulated data.

The draft also discusses interoperability with PIM routers that do not understand the new packet format.

1.1. Conventions used in this document

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

RP: Rendezvous Point

DR: Designated Router

2. Packed Null-Register Capability

A router (DR) can decide to pack multiple Null-Register messages based on the capability received from the RP as part of the PIM Register-Stop. This ensures compatibility with routers that do not support processing of the new format. The capability information can be indicated by the RP via the PIM Register-Stop message sent to the DR. Thus a DR will switch to the new format only when it learns that the RP is capable of handling the PIM Packed Null-Register messages.

Conversely, a DR that does not support the packed format can continue generating the PIM Null-Register as defined in [RFC7761] (Section 4.4). To exchange the capability information in the Register-Stop message, the "Reserved" field can be used to indicate this capability in those Register-Stop messages. One bit of the Reserved field is used to indicate the "packing" capability (P bit). The rest of the bits in the "Reserved" field will be retained for future use.

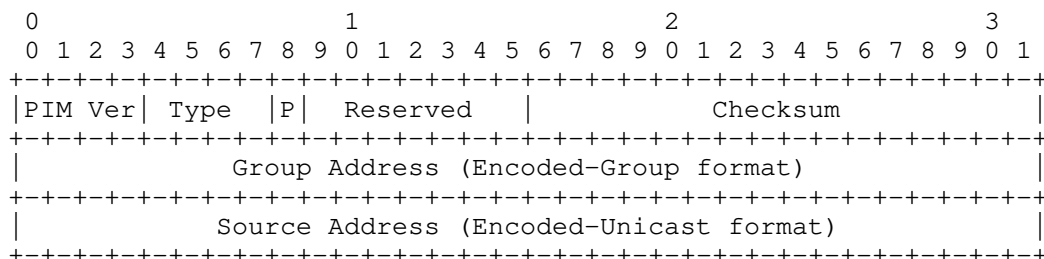


Figure 1: PIM Register-Stop message with capability option

PIM Version, Type, Checksum, Group Address, Source Address:

Same as [RFC7761] (Section 4.9.4)

P:

Capability bit (flag bit 7) used to indicate support for the
Packed Null-Register Capability

3. PIM Packed Null-Register message format

PIM Packed Null-Register message format includes a count to indicate the number of Null-Register records in the message.

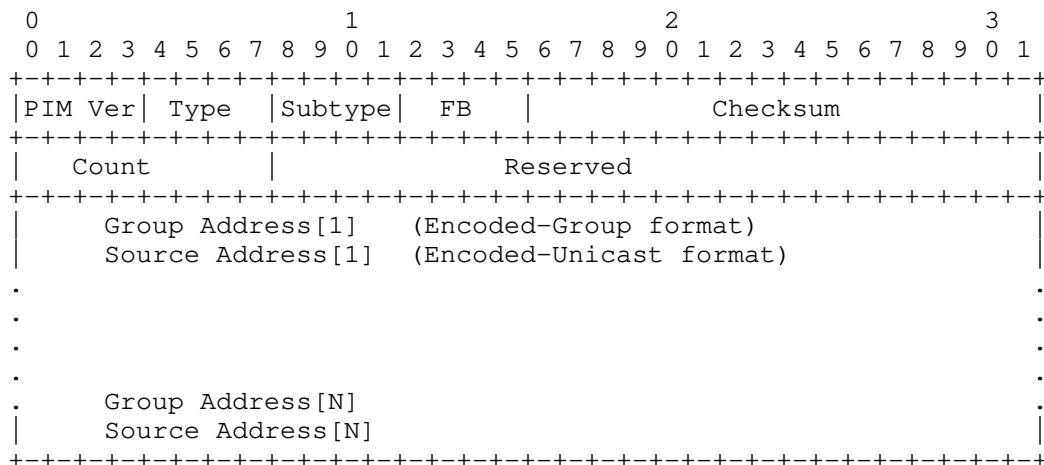


Figure 2: PIM Packed Null-Register message format

PIM Version, Reserved, Checksum:

Same as [RFC7761] (Section 4.9.3)

Type, SubType:

The new packed Null-Register Type and SubType values TBD.
[RFC8736]

Count:

The number of packed Null-Register records. A record consists of a Group Address and Source Address pair.

Group Address, Source Address:

Same as [RFC7761] (Section 4.9.4)

4. PIM Packed Register-Stop message format

The PIM Packed Register-Stop message includes a count to indicate the number of records that are present in the message.

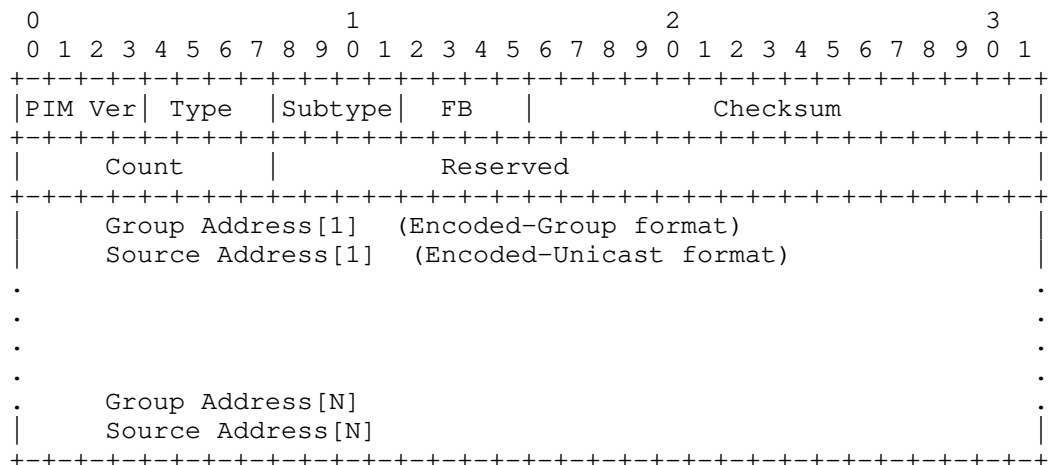


Figure 3: PIM Packed Register-Stop message format

PIM Version, Reserved, Checksum:

Same as [RFC7761] (Section 4.9.4)

Type:

The new Register Stop Type and SubType values TBD

Count:

The number of PIM packed Register-Stop records. A record consists of a Group Address and Source Address pair.

Group Address, Source Address:

Same as [RFC7761] (Section 4.9.4)

5. Protocol operation

The following combinations exist -

1. DR and RP both support the PIM Packed Null-Register and PIM Packed Register-Stop formats:
 - * As specified in [RFC7761], the DR sends PIM Register messages towards the RP when a new source is detected.
 - * An RP supporting this specification MUST set the P-bit in the corresponding Register-Stop messages.
 - * When a Register-Stop message with the P-bit set is received, the DR SHOULD send PIM Packed Null-Register messages (Section 3) to the RP instead of multiple Register messages with the N-bit set [RFC7761].
 - * The RP, after receiving a PIM Packed Null-Register message SHOULD start sending PIM Packed Register-Stop messages (Section 4) to the corresponding DR instead of individual Register-Stop messages.
2. DR supports but RP does not support the PIM Packed Null-Register and PIM Packed Register-Stop formats:
 - * As specified in [RFC7761], DR sends PIM Null-Registers towards the RP.
 - * After receiving DR's PIM Null-Register message, RP sends a normal Register-Stop without any capability information.
 - * DR then sends PIM Null-Registers in the unpacked format [RFC7761].
3. RP supports but DR does not support the PIM Packed Null-Register and PIM Packed Register-Stop formats:

- * As specified in [RFC7761], DR sends the PIM Null-Register towards the RP.
- * After receiving DR's PIM Null-Register message, RP sends a PIM Packed Register-Stop towards the DR that includes capability information.
- * Since DR does not support the new format, it sends PIM Null-Registers in the unpacked format [RFC7761].

6. Operational Considerations

In case the network manager disables the packed capability at the RP, the router should not advertise the capability. However, an implementation MAY choose to still parse any packed registers if they are received. This may be particularly useful in the transitional period after the network manager disables it.

7. PIM Anycast RP Considerations

The PIM Packed Null-Register format should be enabled only if it is supported by all PIM Anycast RP [RFC4610] members in the RP set for the RP address. This consideration applies to PIM Anycast RP with MSDP [RFC3446] as well.

8. PIM RP router version downgrade

Consider a PIM RP router that supports PIM Packed Null-Registers and PIM Packed Register-Stops. When this router downgrades to a software version which does not support PIM Packed Null-Registers and PIM Packed Register-Stops, the DR that sends the PIM Packed Null-Register message will not get a PIM Register-Stop message back from the RP. In such scenarios the DR can send an unpacked PIM Null-Register and check the PIM Register-Stop to see if the capability bit (P-bit) for PIM Packed Null-Register is set or not. If it is not set then the DR will continue sending unpacked PIM Null-Register messages.

9. Fragmentation Considerations

When building a PIM Packed Null-Register message or PIM Packed Register-Stop message, a router should include as many records as possible based on the path MTU towards RP, if path MTU discovery is done. Otherwise, the number of records should be limited by the MTU of the outgoing interface.

10. Security Considerations

General Register messages security considerations from [RFC7761] apply. As mentioned in [RFC7761], PIM Null-Register messages and Register-Stop messages are forwarded by intermediate routers to their destination using normal IP forwarding. Without data origin authentication, an attacker who is located anywhere in the network may be able to forge a Null-Register or Register-Stop message. We next consider the effect of a forgery of each of these messages. By forging a Register message, an attacker can cause the RP to inject forged traffic onto the shared multicast tree.

By forging a Register-Stop message, an attacker can prevent a legitimate DR from registering packets to the RP. This can prevent local hosts on that LAN from sending multicast packets. The above two PIM messages are not changed by intermediate routers and need only be examined by the intended receiver. Thus, these messages can be authenticated end-to-end. Attacks on Register and Register-Stop messages do not apply to a PIM-SSM-only implementation, as these messages are not used in PIM-SSM.

There is another case where a spoofed Register-Stop can be sent to make it appear that is from the RP, and that the RP supports this new packed capability when it does not. This can cause Null-Registers to be sent to an RP that doesn't support this packed format. But standard methods to prevent spoofing should take care of this case. For example, uRPF can be used to filter out packets coming from the outside from addresses that belong to routers inside.

11. IANA Considerations

This document requires the assignment of Capability bit (P-bit), flag bit 7 in the PIM Register-Stop message.

This document requires the assignment of 2 new PIM message types for the PIM Packed Null-Register and PIM Packed Register-Stop.

12. Acknowledgments

The authors would like to thank Stig Venaas, Anish Peter, Zheng Zhang and Umesh Dudani for their helpful comments on the draft.

13. References

13.1. Normative References

- [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>>.
- [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>>.
- [RFC4610] Farinacci, D. and Y. Cai, "Anycast-RP Using Protocol Independent Multicast (PIM)", RFC 4610, DOI 10.17487/RFC4610, August 2006, <<https://www.rfc-editor.org/info/rfc4610>>.
- [RFC8736] Venaas, S. and A. Retana, "PIM Message Type Space Extension and Reserved Bits", RFC 8736, DOI 10.17487/RFC8736, February 2020, <<https://www.rfc-editor.org/info/rfc8736>>.

13.2. Informative References

- [RFC3446] Kim, D., Meyer, D., Kilmer, H., and D. Farinacci, "Anycast Rendezvous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)", RFC 3446, DOI 10.17487/RFC3446, January 2003, <<https://www.rfc-editor.org/info/rfc3446>>.

Authors' Addresses

Vikas Ramesh Kamath
VMware
3401 Hillview Ave
Palo Alto, CA 94304
United States of America

Email: vkamath@vmware.com

Ramakrishnan Chokkanathapuram Sundaram
Cisco Systems, Inc.
Tasman Drive
San Jose, CA 95134
United States of America

Email: ramaksun@cisco.com

Raunak Banthia
Apstra
333 Middlefield Rd STE 200
Menlo Park, CA 94025
United States of America

Email: rbanthia@apstra.com

Ananya Gopal
Cisco Systems, Inc.
Tasman Drive
San Jose, CA 95134
United States of America

Email: ananygop@cisco.com

PIM Working Group
Internet Draft
Intended status: Standards Track
Expires: August 6, 2020

Yisong Liu
China Mobile
M. McBride
T. Eckert
Futurewei
Feb 6, 2020

PIM Assert Message Packing
draft-liu-pim-assert-packing-02

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on August 6, 2020.

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 (<http://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.

Abstract

In PIM-SM shared LAN networks, there is typically more than one upstream router. When duplicate data packets appear on the LAN from different routers, assert packets are sent from these routers to elect a single forwarder. The PIM assert packets are sent periodically to keep the assert state. The PIM assert packet carries information about a single multicast source and group, along with the metric-preference and metric of the route towards the source or RP. This document defines a standard to send and receive multiple multicast source and group information in a single PIM assert packet in a shared network. This can be particularly helpful when there is traffic for a large number of multicast groups.

Table of Contents

1. Introduction	3
1.1. Requirements Language	3
1.2. Terminology	3
2. Use Cases	3
2.1. Enterprise network	4
2.2. Video surveillance	4
2.3. Financial Services	4
2.4. IPTV broadcast Video	4
2.5. Summary	4
3. Solution	5
3.1. PIM Assert Packing Hello Option	5
3.2. PIM Assert Packing Simple Type	5
3.3. PIM Assert Packing Aggregation Type	6
4. Packet Format	6
4.1. PIM Assert Packing Hello Option	6
4.2. PIM Assert Simple Packing Format	7
4.3. PIM Assert Aggregation Packing Format	8
5. IANA Considerations	11
6. Security Considerations	11
7. References	11
7.1. Normative References	11
7.2. Informative References	12
8. Acknowledgments	12
Authors' Addresses	13

1. Introduction

In PIM-SM shared LAN networks, there is typically more than one upstream router. When duplicate data packets appear on the LAN, from different upstream routers, assert packets are sent from these routers to elect a single forwarder according to [RFC7761]. The PIM assert packets are sent periodically to keep the assert state. The PIM assert packet carries information about a single multicast source and group, along with the corresponding metric-preference and metric of the route towards the source or RP.

This document defines a standard to send and receive multiple multicast source and group information in a single PIM assert packet in a shared LAN network. It can efficiently pack multiple PIM assert packets into a single message and reduce the processing pressure of the PIM routers. This can be particularly helpful when there is traffic for a large number of multicast groups.

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

RPF: Reverse Path Forwarding

RP: Rendezvous Point

SPT: Shortest Path Tree

RPT: RP Tree

DR: Designated Router

BDR: Backup Designated Router

2. Use Cases

PIM Assert will happen in many services where multicast is used and not limited to the examples described below.

2.1. Enterprise network

When an Enterprise network is connected through a layer-2 network, the intra-enterprise runs layer-3 PIM multicast. The different sites of the enterprise are equivalent to the PIM connection through the shared LAN network. Depending upon the locations and amount of groups there could be many asserts on the first hop routers.

2.2. Video surveillance

Video surveillance deployments have migrated from analog based systems to IP-based systems oftentimes using multicast. In the shared LAN network deployments, when there are many cameras streaming to many groups there may be issues with many asserts on first hop routers.

2.3. Financial Services

Financial services extensively rely on IP Multicast to deliver stock market data and its derivatives, and current multicast solution PIM is usually deployed. As the number of multicast flows grows, there are many stock data with many groups may result in many PIM asserts on a shared LAN network from publisher to the subscribers.

2.4. IPTV broadcast Video

PIM DR and BDR deployments are often used in host-side network for IPTV broadcast video services. Host-side access network failure scenario may be benefitted by assert packing when many groups are being used. According to [RFC7761] the DR will be elected to forward multicast traffic in the shared access network. When the DR recovers from a failure, the original DR starts to send traffic, and the current DR is still forwarding traffic. In the situation multicast traffic duplication maybe happen in the shared access network and can trigger the assert progress.

2.5. Summary

In the above scenarios, the existence of PIM assert process depends mainly on the network topology. As long as there is a layer 2 network between PIM neighbors, there may be multiple upstream routers, which can cause duplicate multicast traffic to be forwarded and assert process to occur.

Moreover as the multicast services become widely deployed, the number of multicast entries increases, and a large number of assert messages may be sent in a very short period when multicast data packets trigger PIM assert process in the shared LAN networks. The

PIM routers need to process a large number of PIM assert small packets in a very short time. As a result, the device load is very large. The assert packet may not be processed in time or even is discarded, thus extending the time of traffic duplication in the network.

Additionally, future backhaul, or fronthaul, networks may want to connect L3 across an L2 underlay supporting Time Sensitive Networks (TSN). The infrastructure may run DetNet over TSN. These transit L2 LANs would have multiple upstreams and downstreams. This document is taking a proactive approach to prevention of possible future assert issues in these types of environments.

3. Solution

The change to the PIM assert includes two elements: the PIM assert packing hello option and the PIM assert packing method.

There is no change required to the PIM assert state machine. Basically a PIM router can now be the assert winner or loser for multiple packed (S, G)'s in a single assert packet instead of one (S, G) assert at a time. An assert winner is now responsible for forwarding traffic from multiple (S, G)'s out of a particular interface based upon the multiple (S, G)'s packed in a single assert.

3.1. PIM Assert Packing Hello Option

The newly defined Hello Option is used by a router to negotiate the assert packet packing capability. It can only be used when all PIM routers, in the same shared LAN network, support this capability. This document defines two packing methods. One method is a simple merge of the original messages and the other is to extract the common message fields for aggregation.

3.2. PIM Assert Packing Simple Type

In this type of packing, the original assert message body is used as a record. The newly defined assert message can carry multiple assert records and identify the number of records.

This packing method is simply extended from the original assert packet, but, because the multicast service deployment often uses a small number of sources and RPs, there may be a large number of assert records with the same metric preference or route metric field, which would waste the payload of the transmitted message.

3.3. PIM Assert Packing Aggregation Type

When the source or RP addresses, in the actual deployment of the multicast service, are very few, this type of packing will combine the records related to the source address or RP address in the assert message.

* A (S, G) assert only can contain one SPT (S, G) entry, so it can be aggregated according to the same source address, and then all SPT (S, G) entries corresponding to the same source address are merged into one assert record.

* A (*, G) assert may contain a (*, G) entry or a RPT (S, G) entry, and both entry types actually depend on the route to the RP. So it can be aggregated further according to the same RP address, and then all (*, G) and RPT (S, G) entries corresponding to the same RP address are merged into one assert record.

This method can optimize the payload of the transmitted message by merging the same field content, but will add the complexity of the packet encapsulation and parsing.

4. Packet Format

This section describes the format of new PIM messages introduced by this document. The messages follow the same transmission order as the messages defined in [RFC7761].

4.1. PIM Assert Packing Hello Option

```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      OptionType = TBD      |      OptionLength = 1      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Packing_Type   |
+---+---+---+---+---+---+

```

- OptionType: TBD

- OptionLength: 1

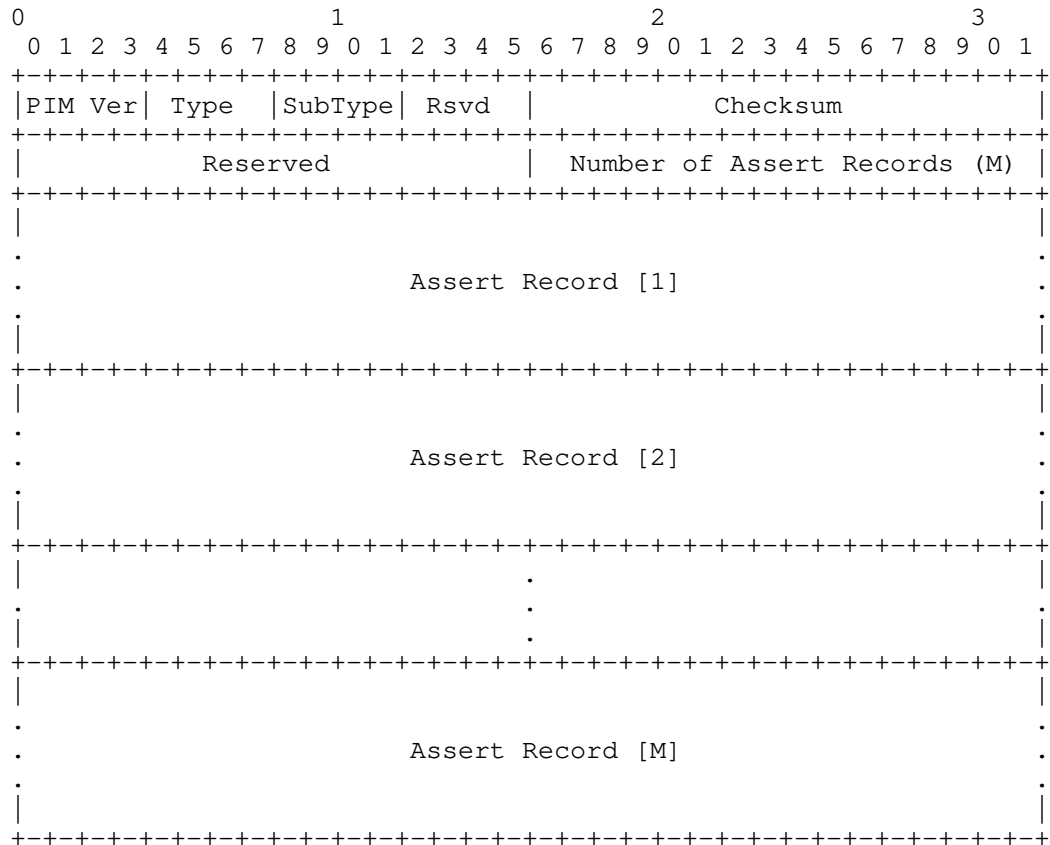
- Packing_Type: The specific packing mode is determined by the value of this field:

1: indicates simple packing type as described in section 2.2

2: indicates aggregating packing type as described in section 2.3

3-255: reserved for future

4.2. PIM Assert Simple Packing Format



PIM Version, Reserved, Checksum

Same as [RFC7761] Section 4.9.6

Type

The new Assert Type and SubType values TBD

Number of Assert Records

The number of packed assert records. A record consists of a single assert message body.

The format of each record is the same as the PIM assert message body of section 4.9.6 in [RFC7761].

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Group Address (Encoded-Group format)               |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Source Address (Encoded-Unicast format)             |
+-----+-----+-----+-----+-----+-----+-----+-----+
|R|               Metric Preference                                 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Metric                                              |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

4.3. PIM Assert Aggregation Packing Format

This method also extends PIM assert packets to carry multiple records. The specific assert packet format is the same as section 4.2, but the records are divided into two types.

The (S, G) assert records are organized by the same source address, and the specific message format is:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Source Address (Encoded-Unicast format)             |
+-----+-----+-----+-----+-----+-----+-----+-----+
|0|               Metric Preference                                 |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Metric                                              |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Number of Groups (N) |               Reserved       |
+-----+-----+-----+-----+-----+-----+-----+-----+
|               Group Address 1 (Encoded-Group format)             |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

```

+-----+
|                                     |
|               Group Address 2 (Encoded-Group format)               |
|                                     |
|                                     |
|                                     |
|               Group Address N (Encoded-Group format)               |
|                                     |
+-----+

```

Source Address, Metric Preference, Metric and Reserved

Same as [RFC7761] Section 4.9.6, but the source address MUST NOT be set to zero.

Number of Groups

The number of group addresses corresponding to the source address field in the (S, G) assert record.

Group Address

Same as [RFC7761] Section 4.9.6, but there are multiple group addresses in the (S, G) assert record

The (*, G) assert records are organized in the same RP address and are divided into two levels of TLVs. The first level is the group record of the same RP address, and the second level is the source record of the same multicast group address, including (*, G) and RPT (S, G), and the specific message format is:

```

      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
+-----+-----+-----+-----+
|                                     |
|               RP Address (Encoded-Unicast format)               |
|                                     |
| 1 |                                     |
|                                     |
|               Metric Preference                                     |
|                                     |
|               Metric                                               |
|                                     |
|   Number of Group Records (0)   |               Reserved         |
|                                     |
+-----+-----+-----+-----+
|                                     |
|                                     |
|                                     |
|               Group Record [1]                                     |
|                                     |
+-----+-----+-----+-----+

```



```

+-----+
|           Source Address 2 (Encoded-Unicast format)           |
+-----+
|           .           |
|           |           |
+-----+
|           Source Address P (Encoded-Unicast format)           |
+-----+

```

Group Address and Reserved

Same as [RFC7761] Section 4.9.6

Number of Sources

The number of source addresses corresponding to the group address field in the group record.

Source Address

Same as [RFC7761] Section 4.9.6, but there are multiple source addresses in the group record.

5. IANA Considerations

This document requests IANA to assign a registry for PIM assert packing Hello Option in the PIM-Hello Options and new PIM assert packet type and subtype. The assignment is requested permanent for IANA when this document is published as an RFC. The string TBD should be replaced by the assigned values accordingly.

6. Security Considerations

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

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.

[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)", RFC 7761, March 2016

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

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

Email: liuyisong@chinamobile.com

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Toerless Eckert
Futurewei

Email: tte+ietf@cs.fau.de

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 7, 2020

M. Sivakumar
Juniper Networks
S. Venaas
Cisco Systems, Inc.
Z. Zhang
ZTE Corporation
March 6, 2020

IGMPv3/MLDv2 Message Extension
draft-venaas-pim-igmp-mld-extension-01

Abstract

IGMP and MLD protocols are extensible, but no extensions have been defined so far. This document provides a well-defined way of extending IGMP and MLD, including a new extension type to distinguish between different extensions.

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 September 7, 2020.

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. Conventions used in this document	2
3. Multicast Listener Query Extension	2
4. Version 2 Multicast Listener Report Extension	4
5. IGMP Membership Query Extension	4
6. IGMP Version 3 Membership Report Extension	5
7. Security Considerations	6
8. IANA Considerations	6
9. References	7
9.1. Normative References	7
9.2. Informative References	7
Authors' Addresses	7

1. Introduction

In this document, we describe a generic method to extend IGMPv3 [RFC3376] and MLDv2 [RFC3810] messages to accommodate information other than what is contained in the current message formats. This is done by introducing an extension-type field in the message formats to indicate the application for which the extension is done. This will be followed by the actual value of the extension.

The extension will be part of additional data as mentioned in [RFC3810] Section 5.1.12 (resp. [RFC3376] Section 4.1.10) for query messages and [RFC3810] Section 5.2.12 (resp. [RFC3376] Section 4.2.11) for report messages.

One such extension is being defined in [I-D.ietf-bier-mld]

2. Conventions used in this document

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.

3. Multicast Listener Query Extension

The MLD query format with extension is shown below

0

1

2

3

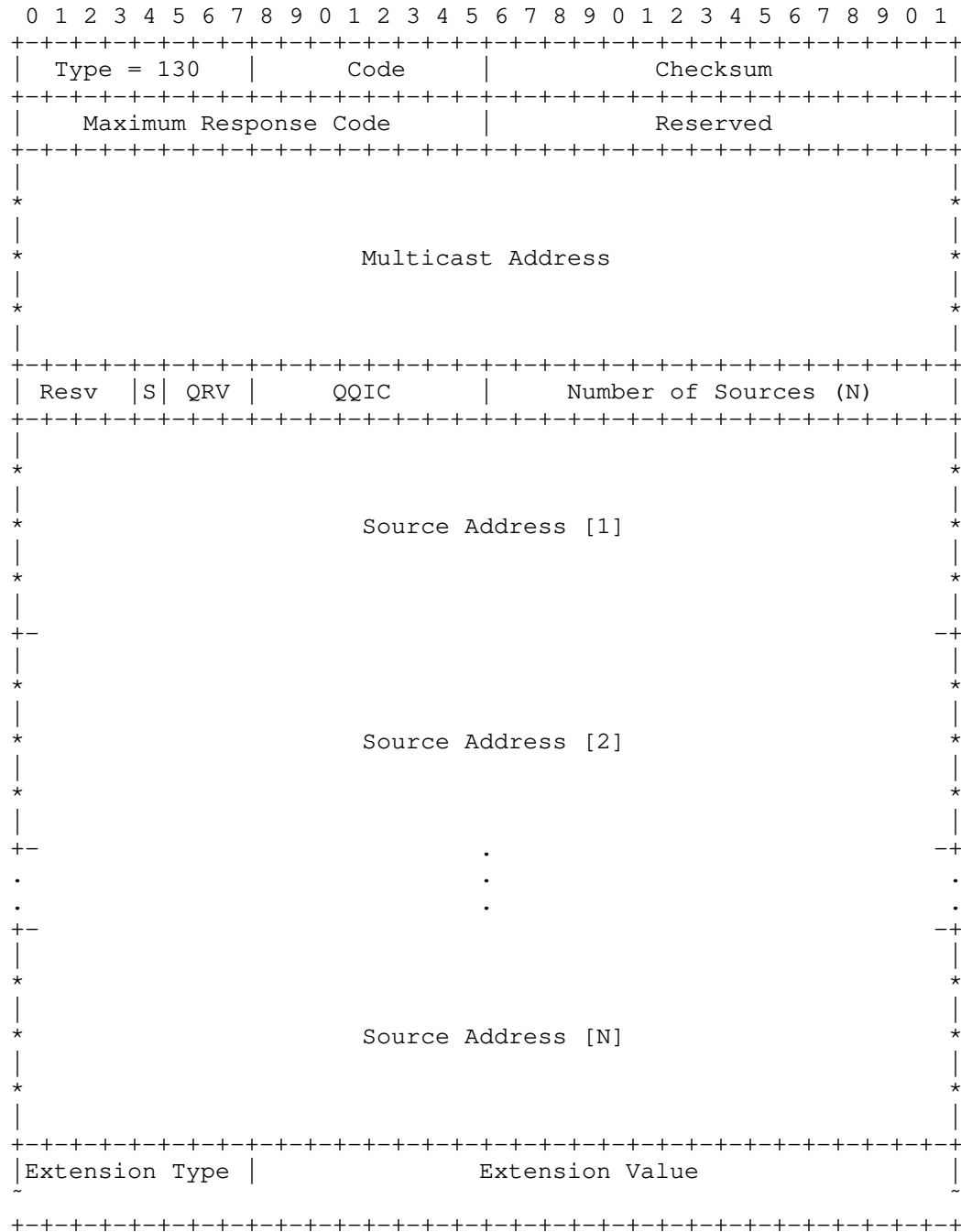


Figure 2: MLD Query Extension

4. Version 2 Multicast Listener Report Extension

The MLD report format with extension is shown below

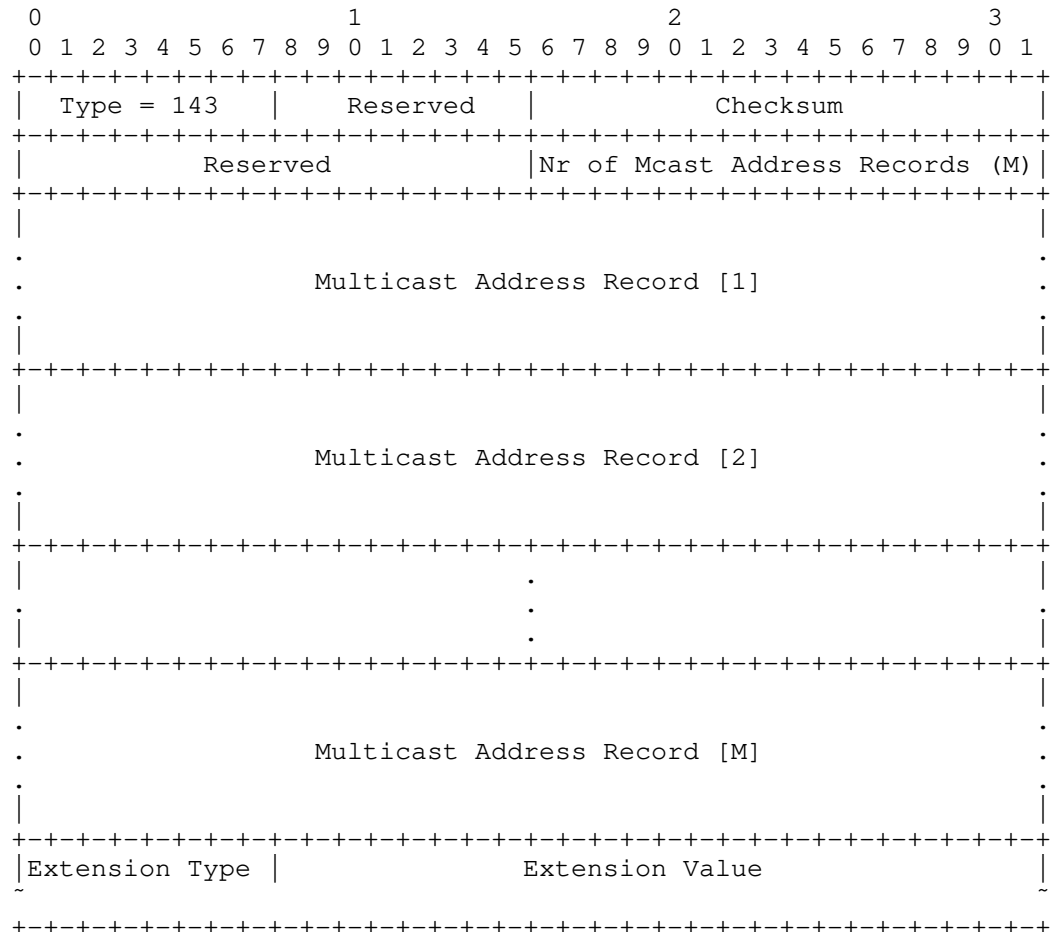


Figure 3: MLD Report Extension

5. IGMP Membership Query Extension

The IGMP query format with the extension is shown below

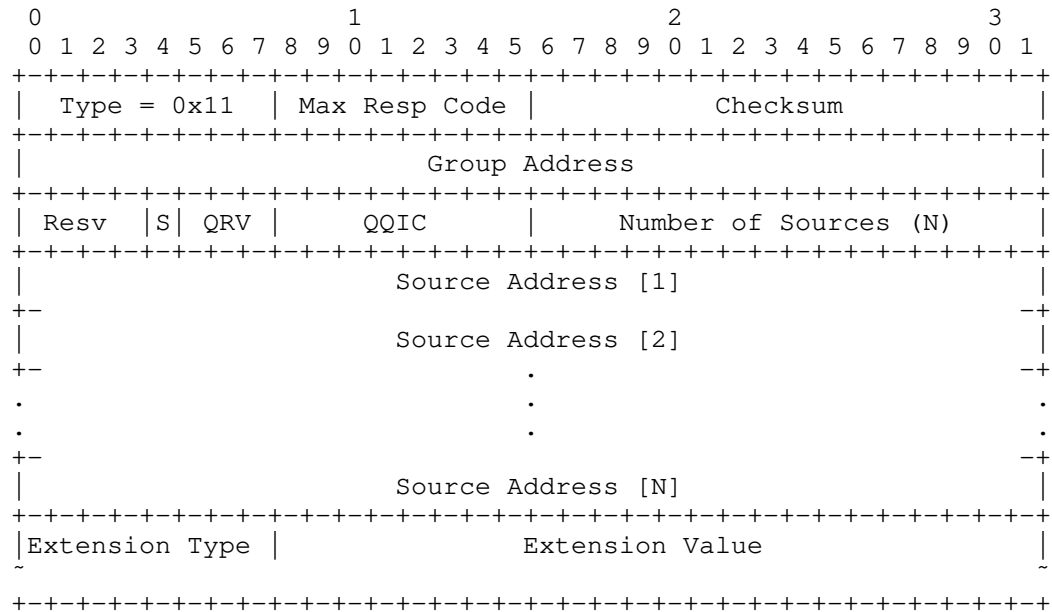


Figure 4: IGMP Query Extension

6. IGMP Version 3 Membership Report Extension

The IGMP report format with the extension is shown below

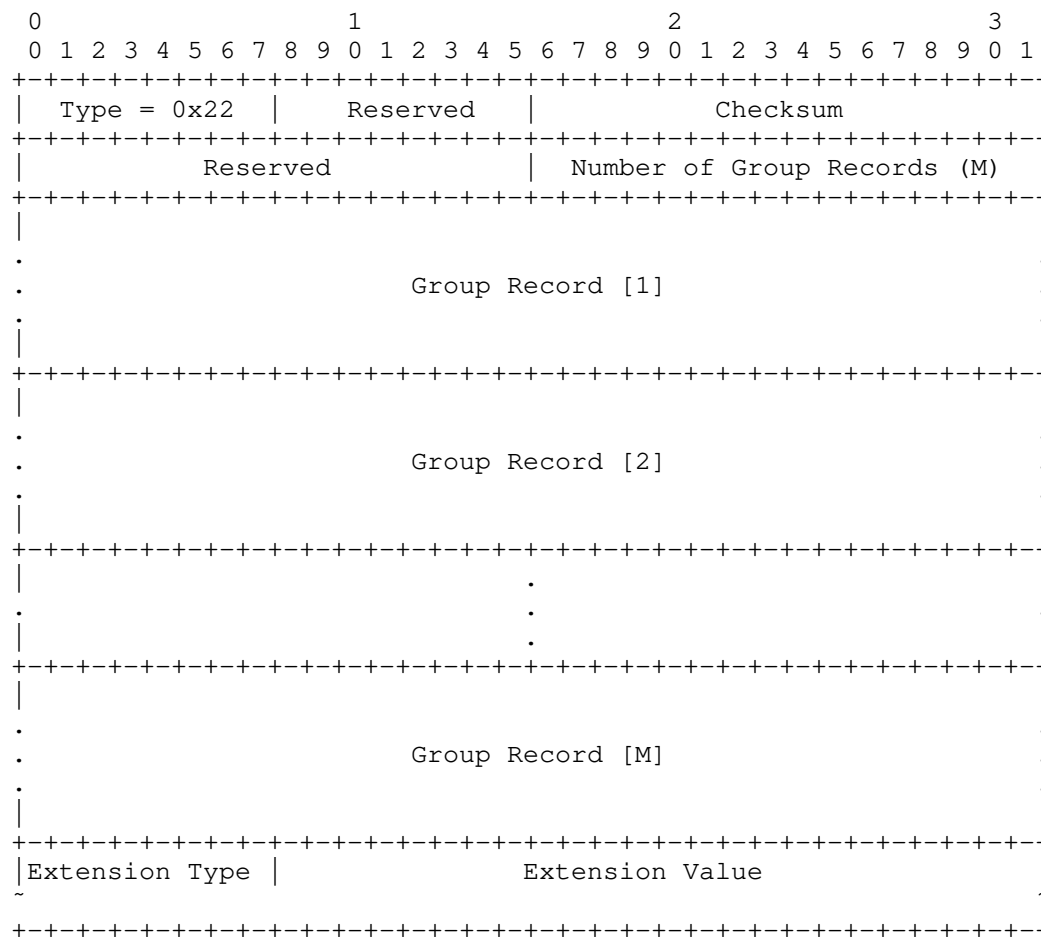


Figure 5: IGMP Report Extension

7. Security Considerations

This document extends MLD (resp. IGMP) message formats. As such, there is no impact on security or changes to the considerations in [RFC3810] and [RFC3376].

8. IANA Considerations

This document requests that IANA creates a new registry for IGMP/MLD extension-types.

9. References

9.1. Normative References

- [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>>.
- [RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, DOI 10.17487/RFC3376, October 2002, <<https://www.rfc-editor.org/info/rfc3376>>.
- [RFC3810] Vida, R., Ed. and L. Costa, Ed., "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, DOI 10.17487/RFC3810, June 2004, <<https://www.rfc-editor.org/info/rfc3810>>.
- [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>>.

9.2. Informative References

- [I-D.ietf-bier-mld] Pfister, P., Wijnands, I., Venaas, S., Wang, C., Zhang, Z., and M. Stenberg, "BIER Ingress Multicast Flow Overlay using Multicast Listener Discovery Protocols", draft-ietf-bier-mld-04 (work in progress), March 2020.

Authors' Addresses

Mahesh Sivakumar
Juniper Networks
64 Butler St
Milpitas CA 95035
USA

Email: sivakumar.mahesh@gmail.com

Stig Venaas
Cisco Systems, Inc.
Tasman Drive
San Jose CA 95134
USA

Email: stig@cisco.com

Zheng (Sandy) Zhang
ZTE Corporation
No. 50 Software Ave, Yuhuatai District
Nanjing 210000
China

Email: zhang.zheng@zte.com.cn