PIM
Internet-Draft
Intended status: Standards Track
Expires: 15 October 2025

Z. Zhang
B. Xu
ZTE Corporation
S. Venaas
Cisco Systems, Inc.
Z. Zhang
Juniper Networks
H. Bidgoli
Nokia

13 April 2025

Multi-Topology in PIM
draft-xz-pim-flex-algo-04

Abstract

PIM usually uses the shortest path computed by routing protocols to build multicast tree. Multi-Topology Routing is a technology to enable service differentiation within an IP network. IGP Flex Algorithm provides a way to compute constraint-based paths over the network. This document defines the PIM message extensions to provide a way to build multicast tree through the specific topology and constraint-based path instead of the shortest path.

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 15 October 2025.

Copyright Notice

Table of Contents

1. Introduction

As described in section 3 in [RFC7761], PIM relies on an underlying topology-gathering protocol to populate the MRIB (Multicast Routing Information Base). Usually the MRIB is the best paths over the network based on the IGP metric. In some cases, alternative paths with low latency or high bandwidth are needed for specific requirements.

Multi-Topology Routing (MTR) [RFC4915] [RFC5120] is a technology to enable service differentiation within an IP network. To support MTR, an IGP advertises multiple, potentially incongruent, IP topologies, and computes topology specific routes.

Flex-Algo [RFC9350] specifies a solution that allows IGPs themselves to compute constraint-based paths over the network. Flex-Algo(FA) can be seen as creating a sub-topology within a topology using algorithm specific constraints and an algorithm specific calculation type. Flex-algo can operate on any topology supported by the IGPs.

Advertisement of IGP Flex-Algo [RFC9350] participation requires a data plane context. Currently the following dataplane contexts have been defined:

* Segment Routing Dataplane [RFC8665] [RFC8667]

* IP Flex Dataplane [RFC9502]

* Soft dataplane [I-D.ginsberg-lsr-flex-soft-dataplane]

This document defines how PIM can utilize a given combination of a topology, an algorithm, and a dataplane to populate an MRIB. In the remainder of this document, we'll refer to this combination as Topology-Algorithm-Dataplane (TAD).

All the routers on a given PIM multicast tree MUST participate in the same TAD.

This document defines the PIM message extensions to provide a way to build a multicast tree using a given TAD instead of simple IGP shortest path.

1.1. Requirements Language

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

2. Terminology

This document uses terminologies defined in [RFC7761], [RFC5384], [RFC5496], [RFC4915], [RFC5120] and [RFC9350].

3. PIM Message extensions

3.1. Group Source Info TAD Sub-TLV

[I-D.ietf-pim-pfm-forwarding-enhancements] defines a 'Group Source Info TLV' for announcing sources that supports Sub-TLVs that can be used to advertise various types of information. This document defines a new Sub-TLV that can be used for carrying the topology ID and the Algorithm value associated with the TAD to be used.

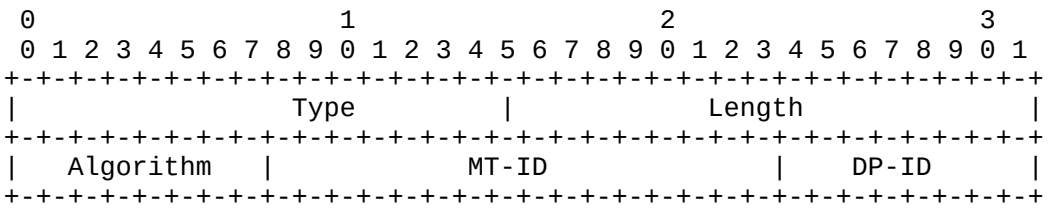

Figure 1

Type: TBD (To be assigned by IANA).

Length: 2-octet. This length is always 4.

Algorithm: A 1-octet value from the IGP Algorithm Types registry under IGP Parameters registry.

MT-ID: A 2-octet field MT-ID (see Section 3.7 of [RFC4915], Section 7 of [RFC5120]) to special the topology. If this field is set to zero, it means the default topology.

DP-ID: A 1-octet field the data plane ID.

MT-ID values are protocol specific. The value advertised therefore has to match an MT-ID value supported by the IGP deployed in the network.

3.2. TAD Attribute Format

[RFC5384] defines a pim Join Attributes are encoded as TLVs into the Encoded-Source Address field of a PIM Join message. This document specifies the TAD Attribute that allows the receiver to select the associated topology or algorithm.

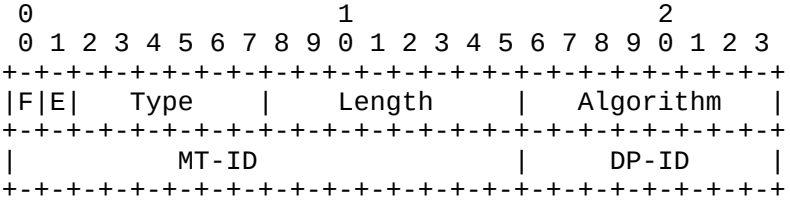

Figure 2

F bit: The Transitive bit. Specifies whether the attribute is transitive or non-transitive. This bit RECOMMENDED set to 1 and the attribute will be transitived.

E bit: End-of-Attributes bit. Specifies whether this attribute is the last. Set to zero if there are more attributes. Set to 1 if this is the last attribute.

Type: TBD (To be assigned by IANA).

Length: 1-octet. This length is always 4.

Algorithm: A 1-octet value from the IGP Algorithm Types registry under IGP Parameters registry.

MT-ID: A 2-octet field MT-ID (see Section 3.7 of [RFC4915], Section 7 of [RFC5120]) to special the topology. If this field is set to zero, it means the default topology.

DP-ID: A 1-octet field the data plane ID.

MT-ID values are protocol specific. The value advertised therefore has to match an MT-ID value supported by the IGP deployed in the network.

4. Specification

When TAD is specified, PIM MUST use the topology, algorithm and data-plane specified.

4.1. Source with TAD Sub-TLV advertisement

PIM Flooding Mechanism and Source Discovery [RFC8364] allows for announcement of active sources.
[I-D.ietf-pim-pfm-forwarding-enhancements] defines a ’Group Source Info TLV’ for announcing sources that allows for Sub-TLVs that can be used for providing various types of information. The TAD Sub-TLV is advertised with the Group Source Info TLV, and flooded in the network. When MTR is not deployed in the network, the MT-ID in the Sub-TLV MUST be set to zero.

The First Hop Router (FHR) advertises the announcing sources carrying the TAD Sub-TLV to the network. All the routers in the network receive the information through PFM function. If two or more FHRs announce same source and group with different TAD because of wrong configurations or other reasons, the LHR SHOULD select the TAD by using the lowest or highest originator address. The highest originator address is preferred.

The PFM function defined in [RFC8364] is not changed.

4.2. J/P message Processing

The LHR PIM router on the receiving side specifies the TAD to the multicast source according to the received TAD Sub-TLV or the local policy. The LHR looks up the local TAD aware routing table and gets the upstream neighbor, then the LHR sends the join message to the upstream neighbor with the specified TAD value set in the TAD attribute. The local configured TAD is the same with the advertisement of LHR usually. In case there is inconsistent, the LHR MUST NOT send the J/P message with TAD attribute. When there is no specific TAD in local policy or configuration on LHR, LHR SHOULD use the TAD received by PFM advertisements if there is.

When a PIM router receives a J/P message with the TAD attribute, the router checks all the received join messages, if all the received join messages carried the TAD value, then it looks up the TAD specific unicast routes and selects the incoming interface and upstream neighbor. And the continual join messages keep carrying the TAD Attribute. When the LHR stops to use the function defined in this document, LHR MUST send the associated prune message. And the continual prune message MUST carried the attribute.

When a PIM router receives the join messages from different neighbors for the same (*,G) or (S,G), in case the router finds that not all the received join messages carried the same TAD value, or unicast routing is unreachable in the TAD aware routing table, the router needs to stop the PIM procedure and sends notification to the network administrator. If the router is FHR, the FHR SHOULD NOT forward the multicast flow until all the received join messages carried the same TAD value.

The TAD attribute SHOULD NOT be used with RPF vector attribute([RFC5496]). In case the TAD attribute is also received with the RPF vector attribute, the router SHOULD ignore one of them according to local policy.

There should be no more than one TAD attribute in an Encoded-Source Address when PIM build a join message. If the PIM router receives a join message with multiple TAD attributes in an Encoded-Source Address, the first one is RECOMMENDED be used.

4.3. Example

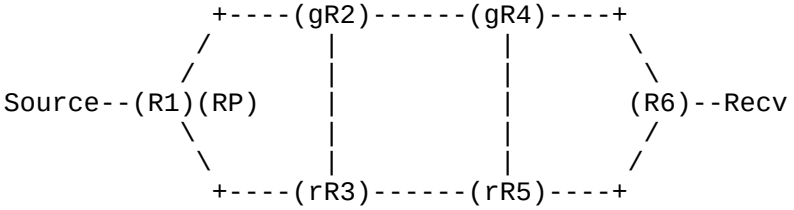

Figure 3

In Figure 3, there is only a default topology in the network. R1 is the source DR and R6 is last-hop DR. Two multicast flows need to be received by the receiver: flow 1 (192.0.2.1/24, 233.252.0.1/32) and flow 2 (198.51.100.1/24, 233.252.0.2/32). The shortest paths to the two sources are the same: R6-R4-R2-R1. But the bandwidth on the path is not enough for the two flows delivery. Packet loss occurs.

The network can be divided into 2 planes by different Flex-Algorithms defined in [RFC9350]. For example, R1/R2/R4/R6 belong to green plane of Algorithm X, and R1/R3/R5/R6 belong to red plane of Algorithm Y.

When the soft dataplane defined in [I-D.ginsberg-lsr-flex-soft-dataplane] is used, the TAD combinations can be TAD 1 "FA=X, MT-ID=0, DP-ID=3" and TAD 2 "FA=Y, MT-ID=0, DP-ID=3". All the routers send the participation of Flex-algo X and Y per [I-D.ginsberg-lsr-flex-soft-dataplane]. The IP prefix routes for the sources of flow 1/2 are advertised in default topology.

R1 sends the PFM messages for flow 1 with TAD 1 sub-TLV and flow 2 with TAD 2 sub-TLV. After receiving the PFM messages, when JoinDesired(192.0.2.1, 233.252.0.1) ([RFC7761]) is TRUE, R6 looks up the local routing by the TAD 1, and gets the upstream router R4 for flow 1. When JoinDesired(198.51.100.1, 233.252.0.2) is TRUE, the process of R6 for TAD 2 is similar, R6 gets the upstream router R5 for flow 2. Then R6 sends the PIM join messages with the TAD 1 to R4 for flow 1 and TAD 2 to R5 for flow 2. All the routers along the path process the join messages in similar way and the multicast trees for flow 1 and flow 2 are built finally.

When the the IP Flex dataplane defined in [RFC9502] is used, the DP-ID field in the TAD combination should be 2. All the routers send the participation of Flex-algo X and Y per [RFC9502]. The IP prefix routes for the sources of flow 1 and 2 are advertised with FA X and Y separately. Similar with the soft dataplane, the multicast trees for flow 1 and flow 2 are built by two different TADs separately.

5. IANA Considerations

5.1. Group Source Info TAD Sub-TLV

IANA is request to assign a new sub-type for "Group Source Info TAD Sub-TLV" in the "PFM Group Source Info Sub-Types" registry.

5.2. TAD Attribute

IANA is request to assign a new sub-type for "TAD Attribute" in the "PIM Join Attribute Types" registry.

5.3. PIM TAD

IANA is request to create a new registry for "Dataplane ID" in the "Interior Gateway Protocol (IGP) Parameters" registry. For now, three values are defined:

* 1: Segment Routing Dataplane [RFC8665] [RFC8667]

* 2: IP Flex Dataplane [RFC9502]

* 3: Soft Dataplane [I-D.ginsberg-lsr-flex-soft-dataplane]

6. Security Considerations

The consideration mentioned in [RFC7761], [I-D.ietf-pim-pfm-forwarding-enhancements] and [I-D.ginsberg-lsr-flex-soft-dataplane] apply to this document.

If PIM routers in the multicast tree select different topology and algorithm based on different local policy, there may be a loop in the network, or the multicast flow cannot be forwarded. Forged router may advertise source and group information with wrong TAD sub-TLV. The network administrator should be careful for the TAD consistency.

7. Acknowledgments

The authors would like to acknowledge Les Ginsberg, Peter Psenak for their input on this document.

8. References

8.1. Normative References

[I-D.ietf-pim-pfm-forwarding-enhancements]
Gopal, A., Venaas, S., and F. Meo, "PIM Flooding Mechanism and Source Discovery Enhancements", Work in Progress, Internet-Draft, draft-ietf-pim-pfm-forwarding-enhancements-01, 3 November 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-pim-pfm-forwarding-enhancements-01>.

[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>.

[RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, DOI 10.17487/RFC4915, June 2007, <https://www.rfc-editor.org/info/rfc4915>.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <https://www.rfc-editor.org/info/rfc5120>.

[RFC5384] Boers, A., Wijnands, I., and E. Rosen, "The Protocol Independent Multicast (PIM) Join Attribute Format", RFC 5384, DOI 10.17487/RFC5384, November 2008, <https://www.rfc-editor.org/info/rfc5384>.

[RFC5496] Wijnands, IJ., Boers, A., and E. Rosen, "The Reverse Path Forwarding (RPF) Vector TLV", RFC 5496, DOI 10.17487/RFC5496, March 2009, <https://www.rfc-editor.org/info/rfc5496>.

[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>.

[RFC9350] Psenak, P., Ed., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", RFC 9350, DOI 10.17487/RFC9350, February 2023, <https://www.rfc-editor.org/info/rfc9350>.

8.2. Informative References

[I-D.ginsberg-lsr-flex-soft-dataplane]
Ginsberg, L., Psenak, P., and Z. Zhang, "IGP Flex Soft Dataplane", Work in Progress, Internet-Draft, draft-ginsberg-lsr-flex-soft-dataplane-01, 13 April 2025, <https://datatracker.ietf.org/doc/html/draft-ginsberg-lsr-flex-soft-dataplane-01>.

[RFC8364] Wijnands, IJ., Venaas, S., Brig, M., and A. Jonasson, "PIM Flooding Mechanism (PFM) and Source Discovery (SD)", RFC 8364, DOI 10.17487/RFC8364, March 2018, <https://www.rfc-editor.org/info/rfc8364>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

[RFC9502] Britto, W., Hegde, S., Kaneriya, P., Shetty, R., Bonica, R., and P. Psenak, "IGP Flexible Algorithm in IP Networks", RFC 9502, DOI 10.17487/RFC9502, November 2023, <https://www.rfc-editor.org/info/rfc9502>.

Authors' Addresses

Zheng Zhang
ZTE Corporation
China
Email: zhang.zheng@zte.com.cn

Benchong Xu
ZTE Corporation
China
Email: xu.benchong@zte.com.cn

Stig Venaas
Cisco Systems, Inc.
Tasman Drive
San Jose, CA 95134,
United States of America

Email: stig@cisco.com

Zhaohui Zhang
Juniper Networks
Boston,
United States of America
Email: zzhang@juniper.net

Hooman Bidgoli
Nokia
Ottawa
Canada
Email: hooman.bidgoli@nokia.com