mboned
Internet-Draft
Intended status: Standards Track
Expires: 3 September 2026

Z. Zhang
L. Giuliano
HPE
B. Tarno
Disney
C. Lenart
Verizon

2 March 2026

Dynamic Internet Multicast Tunneling
draft-zzhang-mboned-dynamic-internet-mcast-tunnel-00

Abstract

This document specifies a mechanism to facilitate widespread multicast connectivity over the Global Internet via dynamic tunneling, enabling many different multicast islands to be connected by tunnels between both PIM routers and AMT gateways/relays.

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 3 September 2026.

Copyright Notice

Table of Contents

1. Introduction

IP Multicast requires every Layer 3 hop between source and receivers to be multicast-enabled. This requirement has always been a significant barrier to deployment of multicast over the Global Internet as it essentially requires enabling a multicast routing protocol on every interface on every router and firewall between all hosts to achieve ubiquitous availability.

To overcome this hurdle, network overlays (or tunnels) have been used to connect multicast-enabled routers/hosts that are separated by unicast-only parts of the network. For example, the original Multicast Backbone (MBONE) was constructed in the 1990s as a network of multicast-enabled devices connected together by GRE tunnels across the Internet. But these GRE tunnels were onerous to manage as they require static configuration and coordination on both ends, plus unicast routing protocols like BGP, ISIS and OSPF to be run through the tunnels so that Reverse-Path Forwarding (RPF) would operate properly. Subsequent efforts in the early 2000s focused on native multicast deployment across the Internet, but these efforts inevitably fell short of global ubiquity.

Automatic Multicast Tunneling (AMT) [RFC7450] was later created as a dynamic tunneling mechanism to overcome the operational shortcomings of static tunneling protocols like GRE. AMT was initially motivated to solve the "last-mile" problem, where receivers on unicast-only networks (AMT gateways) could dynamically build tunnels to devices at the edge of multicast-enabled parts of the network (AMT Relays) and receive multicast content without any dependencies on their local service provider. AMT and SSM combined together to form TreeDN [RFC9706], a tree-based CDN architecture that addressed the operational challenges of multicast over the Global Internet. TreeDN used SSM as a simplified deployment option for efficiently delivering multicast content from the source to the border of the multicast-enabled domain, with AMT providing availability in the "last-mile" by delivering this content to receivers on unicast-only networks.

Over time, there has been a growing interest in using AMT in the "middle-mile" to connect multicast "islands" that are separated by unicast-only transit networks. Strictly speaking, AMT has always had the ability to build router-router tunnels, in addition to router-host tunnels. But because no unicast-routing protocols can run through the AMT tunnels (IGMP is the only control plane protocol that can run through an AMT tunnel), determining source reachability (RPF) is a problem when there are potentially many different AMT relays to choose from, with no obvious way for a transit router acting as AMT Gateway to know which AMT relay has multicast connectivity to a particular source. DRIAD [RFC8777] was proposed to address the AMT Relay discovery problem, but this solution has dependencies on the receiver’ s local service provider- essentially assuming the "last-mile" supports multicast (and DRIAD). Further, DRIAD is more suitable for hosts, not core routers, acting as AMT GWs.

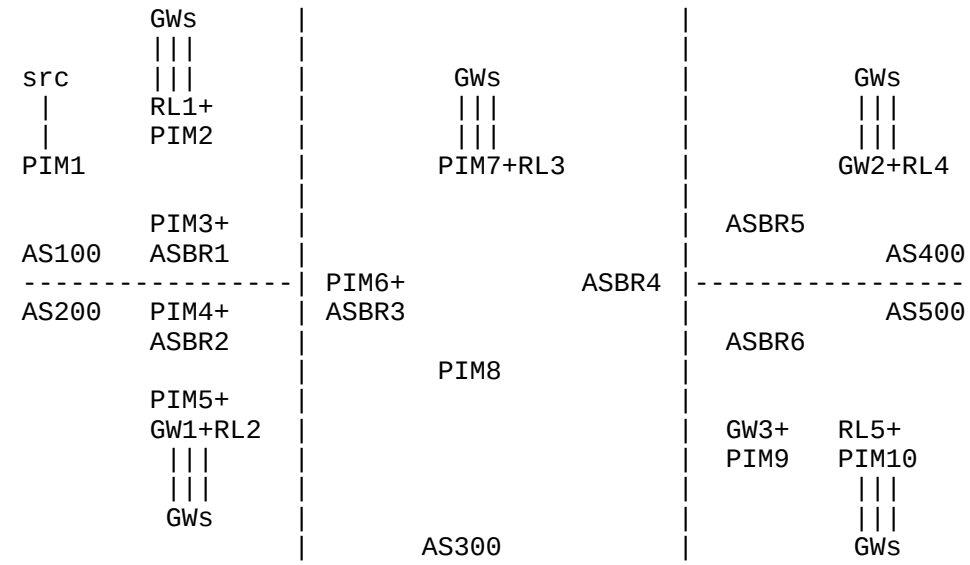
In addition, for the multicast routers running PIM-SSM [RFC7761] over the tunnels between them, it is desired to make the tunnel management automatic, and remove the need for exchanging RPF routes over the tunnels.

This document specifies a mechanism for extending the TreeDN architecture to facilitate widespread multicast connectivity over the Global Internet via dynamic tunneling in the "middle-mile", enabling many different multicast islands to be connected by tunnels between both PIM routers and AMT gateways/relays.

This solution provides the benefits of dynamically created tunnels (i.e., no manual configuration on the tunnel endpoints, as was needed with GRE) while providing reachability info regarding which sources could be reached behind which routers/relays.

2. Mode of Operation

Consider the following topology with five ASes:



GW: AMT Gateway RL: AMT Relay

* All devices with + in the title are single routers serving multiple functions. So PIM5+GW1+RL2 is a PIM router that also functions as both AMT GW and AMT Relay.

* The source is connected to router PIM1 in AS100. There are several local receivers connected as AMT GWs off RL1+PIM2, which is a PIM neighbor of PIM1.

* In AS200, PIM4+ASBR2 is a PIM neighbor of PIM3+ASBR1 in AS100. There are several local receivers connected as AMT GWs to PIM5+GW1+RL2, which runs both PIM and AMT GW (it could receive traffic from either PIM4+ASBR2 natively or from RL1-PIM2 via AMT or from both).

* In AS300, PIM6+ASBR3 runs PIM but ASBR4 does not. PIM7+RL3 has receivers connected as AMT GWs.

* Neither ASBR5 nor ASBR6 runs PIM. In AS400, GW2+RL4, has several local receivers connected as AMT GWs, and joins upstream as an AMT GW connected to PIM7+RL3 in AS300. GW2+RL4 serves as an IGMP proxy since it sends IGMP reports upstream based on the IGMP reports/PIM joins it receives from hosts/routers downstream.

* In AS500, GW3+PIM9 receives traffic from either PIM7+RL3 via AMT (PIM->IGMP proxy) or from PIM8 via a dynamic PIM tunnel. RL5+PIM10 receives traffic from GW3+PIM9, and forwards that traffic to local receivers connected as AMT GWs.

When PIM3+ASBR1 advertises the route for the source to PIM4+ASBR2 and PIM6+ASBR3, it includes an Upstream Multicast Hop (UMH) Extended Community (EC) that encodes RL1+PIM2' s address as an AMT relay, so that downstream routers know that RL1+PIM2 can be used as the AMT relay for multicast traffic from the source prefix.

Suppose PIM4+ASBR2 re-advertises the source's BGP route to PIM5+GW1+RL2 with the EC attached. When PIM5+GW1+RL2 receives IGMP joins from its own downstream AMT GWs, it does the RPF lookup and finds the route with the UMH EC specifying the RL1+PIM2 as the AMT relay. The route also resolves to the next hop router PIM4+ASBR2, so PIM5+GW1+RL2 can choose to pull traffic either via PIM from PIM4+ASBR2, or via AMT from RL1+PIM2. In the latter case, PIM5+GW1+RL2 is an AMT GW of RL1+PIM2.

When PIM7+RL3 in AS300 receives IGMP joins from its own downstream AMT GWs, since it is not an AMT GW-capable router, it can only receive traffic natively from PIM6+ASBR3.

Notice that ASBR4/5/6 do not run PIM in this example. Suppose that when ASBR4 re-advertises the route for the source to ASBR5/6, it keeps the UMH EC unchanged or changes it to encode PIM7+RL3's address, and add another UMH EC that encodes PIM8's address as a PIM router.

Suppose ASBR5/6 re-advertises the route to GW2+RL4 and RL5+PIM10, respectively, keeping the UMH EC. GW2+RL4 will serve as IGMP proxy to PIM7+RL3 via AMT, while RL5+PIM10 sends a PIM join toward GW3+PIM9. Either RL5+PIM10 or GW3+PIM9 may directly send PIM joins to PIM8 over a dynamic tunnel, using the PIM-lite mode [RFC9739] where a PIM adjacency is not required.

There are a few key points to note here:

* A router that receives IGMP joins (via AMT or natively) or PIM joins needs to either have a direct upstream PIM neighbor toward the source/ UMH, or be able to establish an AMT or PIM tunnel to a remote UMH.

* The remote UMHs for a multicast source may be learned from the UMH EC attached to the BGP routes, may be provisioned locally, or may be learned from the PIM RPF Vector attribute [RFC5496] in PIM joins from the downstream.

3. Specification

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.

3.1. Upstream Multicast Hop Extended Community (UMH EC)

The UMH EC is an IPv4 Address Specific Extended Community with a sub-type TBD1 (Section 6) to be assigned by IANA, or an IPv6 Address Specific Extended Community with a type TBD2 (Section 6) to be assigned by IANA.

The Global Admin Field of the EC encodes the address of the UMH. The lower 4-bit of the Local Admin Field encodes the type of the UMH. The next 4-bit encodes the preference of the UMH, and the remaining upper 8-bit of the Local Admin Field is reserved and MUST be set to 0 when sending and MUST be ignored when receiving.

3.2. Procedures

When an ASBR advertises routes to its internal/external peers, it MAY keep the existing UMH ECs unchanged, or remove some/all of them, and/or add its own UMH ECs based local policies. The added UMH ECs MUST be for PIM routers or AMT relays in the ASBR’s own AS that accept joins from AMT GWs or PIM routers over tunnels. When multiple UMH ECs are added, the preference bits SHOULD be set to indicate the preference. The higher the number, the higher preference.

A PIM router or AMT Relay may learn remote UMHs by receiving BGP routes with the UMH ECs or from local provisioning. It may choose to send PIM/IGMP joins over a tunnel to a remote UMH, or tunnels to more UMHs for redundancy purposes as specified in [RFC7431] for Multicast Only Fast ReRoute (MoFRR), or to an immediate upstream PIM router based on RPF lookup and local policies. In the latter case, PIM RPF Vector [RFC5496] may be used so that routers along the path only need to do a RPF check toward the remote UMH instead of the source, so that they do not need to learn the routes to the source.

When there are multiple UMH ECs in the route to the source, the preference order SHOULD be determined as follows:

* If the route to a UMH address has the shortest AS_PATH, the UMH is most preferred.

* For UMHes whose routes have the same length of AS_PATH, the one with the higher preference value in the UMH EC is preferred.

With the same AS_PATH length and preference, it does not matter which UMH is preferred. Local policy MAY also override the preference, and the ultimate choice only has local significance.

When AMT tunnels to the upstream are used, it behaves as an AMT GW (in addition to PIM router or AMT relay), and IGMP proxy (from downstream IGMP or PIM joins) procedures are followed, including the procedures in [I-D.ietf-pim-multipath-igmpmldproxy].

While AMT defines handshake procedures for the tunnel establishment, there are no specified procedures to establish the tunnels between remote PIM neighbors, which operate in PIM-lite mode [RFC9739] between them. A downstream PIM router can use any mechanism to tunnel PIM joins to an upstream PIM router - UDP, GRE, MPLS, or BIER [I-D.ietf-bier-pim-signaling]. The upstream PIM router can use any mechanism to tunnel multicast data to remote downstream neighbors. The only requirement is that downstream routers MUST be able to associate incoming tunneled multicast data with logical interfaces for RPF purposes. The PIM joins sent over the tunnel may carry a PIM Join Attribute [RFC5384] to indicate the desired tunnel, which may also help the downstream router associating the incoming traffic with the logical interface. Details will be specified in a future revision of this document.

4. Operational Considerations

The announcement of UMHs via BGP is like dynamic unicast routing and the provisioning of UMHs on a router is like static unicast routing. Care must be taken that the dynamic and static routing are consistent to prevent multicast routing loops.

An operator should be cautious when its ASBRs add the UMH ECs when re-advertising routes externally, because the announced UMHs may attract many external AMT/PIM joins. AMT has security measures for access control and protection against resource exhaustion, and many PIM implementations also have policies to allow/deny joins for access control. It is a reasonable practice to accept remote PIM/IGMP joins from only neighboring or specifically allowed ASes, and with additional fine grain control by policies.

5. Security Considerations

Routers that originate or change the UMH EC MUST ensure that the UMHs have multicast connectivity to the multicast source, whether natively or through tunnels. Rogue routers, or those with no multicast connectivity to the source, could create multicast blackholes by advertising multicast reachability to the source via the UMH EC. That said, this risk is not much different than what can occur with any other reachability information that BGP advertises (eg, unicast routing).

6. IANA Considerations

This document requests IANA to assign an Extended Community sub-type of value TBD1 for Upstream Multicast Hop from the Transitive IPv4-Address-Specific Extended Community Sub-Types registry.

This document requests IANA to assign an IPv6-Address-Specific Extended Community type of value TBD2 from the Transitive IPv6-Address-Specific Extended Community Types registry.

This document requests IANA to create a Upstream Multicast Hop Types registry with the following initial allocations:

Value	Description	Reference
0	Reserved	This Document
1	PIM	This Document
2	AMT Relay	This Document
3-15	Unallocated	This Document

The registration procedure is "Specification Required".

7. Acknowledgments

The solution described in this document was inspired by earlier discussions with the following people about the desire for using AMT in the "middle mile": Jake Holland, Erik Herz, Guillaume Bichot, Omar Ramadan and Marc Fiuczynski.

8. References

8.1. Normative References

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

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

[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>.

[RFC9739] Bidgoli, H., Ed., Venaas, S., Mishra, M., Zhang, Z., and M. McBride, "Protocol Independent Multicast Light (PIM Light)", RFC 9739, DOI 10.17487/RFC9739, March 2025, <https://www.rfc-editor.org/info/rfc9739>.

[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>.

8.2. Informative References

[RFC9706] Giuliano, L., Lenart, C., and R. Adam, "TreeDN: Tree-Based Content Delivery Network (CDN) for Live Streaming to Mass Audiences", RFC 9706, DOI 10.17487/RFC9706, January 2025, <https://www.rfc-editor.org/info/rfc9706>.

[RFC8777] Holland, J., "DNS Reverse IP Automatic Multicast Tunneling (AMT) Discovery", RFC 8777, DOI 10.17487/RFC8777, April 2020, <https://www.rfc-editor.org/info/rfc8777>.

[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>.

[RFC7431] Karan, A., Filsfils, C., Wijnands, IJ., Ed., and B. Decraene, "Multicast-Only Fast Reroute", RFC 7431, DOI 10.17487/RFC7431, August 2015, <https://www.rfc-editor.org/info/rfc7431>.

[I-D.ietf-pim-multipath-igmpmldproxy]
Asaeda, H. and L. M. Contreras, "Multipath Support for IGMP/MLD Proxy", Work in Progress, Internet-Draft, draft-ietf-pim-multipath-igmpmldproxy-03, 20 October 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-pim-multipath-igmpmldproxy-03>.

[I-D.ietf-bier-pim-signaling]
Bidgoli, H., Xu, F., Kotalwar, J., Wijnands, I., Mishra, M. P., and Z. J. Zhang, "PIM Signaling Through BIER Core", Work in Progress, Internet-Draft, draft-ietf-bier-pim-signaling-13, 3 March 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-bier-pim-signaling-13>.

Authors' Addresses

Zhaohui Zhang
HPE
Email: zhaohui.zhang@hpe.com

Lenny Giuliano
HPE
Email: leonard.giuliano@hpe.com

Brad Tarno
Disney
Email: brad.tarno@disney.com

Christopher Lenart
Verizon
Email: chris.lenart@verizon.com