Workgroup: MBONED
Internet-Draft:

draft-ietf-mboned-multicast-telemetry-08

Published: 6 October 2023

Intended Status: Standards Track

Expires: 8 April 2024

Authors: H. Song M. McBride

Futurewei Technologies Futurewei Technologies

G. Mirsky G. Mishra H. Asaeda

Ericsson Verizon Inc. NICT

T. Zhou

Huawei Technologies

Multicast On-path Telemetry using IOAM

Abstract

This document specifies the requirements of on-path telemetry for multicast traffic using In-situ OAM. While In-situ OAM is advantageous for multicast traffic telemetry, some unique challenges are present. This document provides the solutions based on the In-situ OAM trace option and direct export option to support the telemetry data correlation and the multicast tree reconstruction without incurring data redundancy.

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.

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 8 April 2024.

Copyright Notice

Copyright (c) 2023 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. Requirements for Multicast Traffic Telemetry
- 3. <u>Issues of Existing Techniques</u>
- 4. <u>Modifications to Existing Solutions</u>
 - 4.1. Per-hop postcard using IOAM DEX
 - 4.2. Per-section postcard for IOAM Trace
- 5. Application Considerations for Multicast Protocols
 - 5.1. Mtrace verson 2
 - 5.2. Application in PIM
 - 5.3. Application of MVPN X-PMSI Tunnel Encapsulation Attribute
 - 5.4. Application in BIER
- 6. Security Considerations
- 7. IANA Considerations
- 8. Acknowledgments
- 9. References
 - 9.1. Normative References
 - 9.2. Informative References

<u>Authors' Addresses</u>

1. Introduction

Multicast has many use cases. For example, it can be used by residential broadband customers across operator networks, private MPLS customers, and internal customers within corporate intranet. Multicast provides real time interactive online meetings or podcasts, IPTV, and financial markets real-time data, which all have a reliance on UDP's unreliable transport. End-to-end QOS, therefore, should be a critical component of multicast deployment in order to provide a good end user experience. In multicast real-time media streaming, loss of a single packet containing a reference frame can result in the inability of thousands of receivers to decode a whole sequence of packets called Group-of-Picture, introducing black

picture for periods of a few seconds. Unexpected long delay in propagation of a packet in such real-time media streaming may equally result in the packet not being received and create the same results. Multicast packet drops and delay can therefore severely affect the application performance and user experience.

It is important to monitor the performance of the multicast traffic. New on-path telemetry techniques such as In-situ-OAM (IOAM)

[RFC9197], IOAM Direct Export (DEX) [RFC9326] IOAM Marking-based

Postcard (PBT-M) [I-D.song-ippm-postcard-based-telemetry], and

Hybrid Two-Step (HTS) [I-D.ietf-ippm-hybrid-two-step] are useful and complementary to the existing active OAM performance monitoring methods (e.g., ICMP ping [RFC0792]), provide promising means to directly monitor the network experience of multicast traffic. However, multicast traffic has some unique characteristics which pose some challenges on applying such techniques in an efficient way.

The IP multicast packet data for a particular (S, G) state is identical from one branch to another on its way to multiple receivers. When adding IOAM trace data to multicast packets, each replicated packet would keep the telemetry data for its entire forwarding path. Since the replicated packets all share common path segments, redundant data will be collected for the same original multicast packet. Such redundancy consumes extra network bandwidth unnecessarily. For a large multicast tree, such redundancy is considerable. Alternatively, it could be more efficient to collect the telemetry data using solutions such as IOAM DEX to eliminate the data redundancy. However, IOAM DEX lacks a branch identifier, making telemetry data correlation and multicast-tree reconstruction difficult.

This draft provides two solutions to the IOAM data redundancy problem based on the IOAM standards. The requirements for multicast traffic telemetry are discussed along with the issues of the existing on-path telemetry techniques. We propose modifications to make these techniques adapt to multicast in order for the original multicast tree to be correctly reconstructed while eliminating redundant data.

2. Requirements for Multicast Traffic Telemetry

Multicast traffic is forwarded through a multicast tree. With PIM and P2MP, the forwarding tree is established and maintained by the multicast routing protocol. With BIER, no state is created in the network to establish a forwarding tree; instead, a bier header provides the necessary information for each packet to know the egress points. Multicast packets are only replicated at each tree branch fork node for efficiency.

There are several requirements for multicast traffic telemetry, a few of which are:

- *Reconstruct and visualize the multicast tree through data plane monitoring.
- *Gather the multicast packet delay and jitter performance on each path.
- *Find the multicast packet drop location and reason.
- *Gather the VPN state and tunnel information in case of P2MP multicast.

In order to meet these requirements, we need the ability to directly monitor the multicast traffic and derive data from the multicast packets. The conventional OAM mechanisms, such as multicast ping [RFC6450] and trace [RFC8487], are not sufficient to meet these requirements.

3. Issues of Existing Techniques

On-path Telemetry techniques that directly retrieve data from multicast traffic's live network experience are ideal for addressing the aforementioned requirements. The representative techniques include In-situ OAM (IOAM) <a href="In-situ OAM OIOAM) <a href="In-situ OA

Multicast packets are replicated at each branch fork node in the corresponding multicast tree. Therefore, there are multiple copies of the original multicast packet in the network.

If the IOAM trace option is used for on-path data collection, the partial trace data will also be replicated into the packet copy for each branch. The end result is that, at the multicast tree leaves, each copy of the multicast packet has a complete trace. Most of the data (except data from the last leaf branch) appear in multiple copies while only one copy is sufficient. Data redundancy introduces unnecessary header overhead, wastes network bandwidth, and complicates the data processing. The larger the multicast tree, or the longer the multicast path, the more severe the redundancy problem becomes.

The postcard-based solutions (e.g., IOAM DEX), can be used to eliminate such data redundancy, because each node on the tree only sends a postcard covering local data. However, they cannot track and correlate the tree branches properly due to the lack of branching information, so they can bring confusion about the multicast tree

topology. For example, in a multicast tree, Node A has two branches, one to Node B and the other to node C; further, Node B leads to Node D and Node C leads to Node E. When applying postcard-based methods, one cannot tell whether or not Node D(E) is the next hop of Node B(C) from the received postcards alone, unless one correlates the exporting nodes with knowledge about the tree collected by other means (e.g., mtrace). Such correlation is undesirable because it introduces extra work and complexity.

The fundamental reason for this problem is that there is not an identifier (either implicit or explicit) to correlate the data on each branch.

4. Modifications to Existing Solutions

We provide two solutions to address the above issues. One is based on IOAM DEX and requires an extension to the instruction header of the IOAM DEX Option. The second solution combines the IOAM trace option and postcards for redundancy removal.

4.1. Per-hop postcard using IOAM DEX

One way to mitigate the postcard-based telemetry's tree tracking weakness is to augment it with a branch identifier field. Note that this works for the IOAM DEX option but not for PBT-M because the IOAM DEX option has an instruction header which can be used to hold the branch identifier. To make the branch identifier globally unique, the branch fork node ID plus an index is used. For example, Node A has two branches: one to Node B and the other to Node C. Node A will use [A, 0] as the branch identifier for the branch to B, and [A, 1] for the branch to C. The identifier is carried with the multicast packet until the next branch fork node. Each node MUST export the branch identifier in the received IOAM DEX header in the postcards it sends. The branch identifier, along with the other fields such as flow ID and sequence number, is sufficient for the data collector to reconstruct the topology of the multicast tree.

<u>Figure 1</u> shows an example of this solution. "P" stands for the postcard packet. The square brackets contains the branch identifier. The curly brace contains the telemetry data about a specific node.

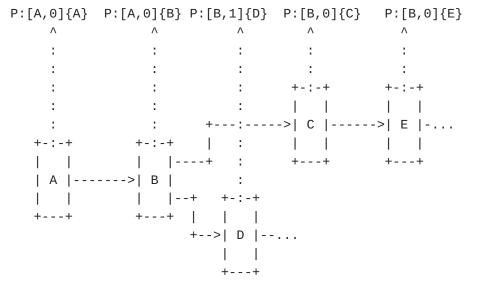


Figure 1: Per-hop Postcard

Each branch fork node needs to generate a unique branch identifier (i.e., branch ID) for each branch in its multicast tree instance and include it in the IOAM DEX option header. The branch ID remains unchanged until the next branch fork node. The branch ID contains two parts: the branch fork node ID and an interface index.

Conforming to the node ID specification in $\underline{\text{IOAM}}$ [RFC9197], the node ID is a 3-octet unsigned integer. The interface index is a two-octet unsigned integer. As shown in $\underline{\text{Figure 2}}$, the branch ID consumes 8 octets in total. The three unused octets MUST be set to 0.

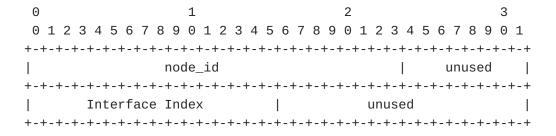


Figure 2: Multicast Branch ID format

Figure 3 shows that the branch ID is carried as an optional field after the flow ID and sequence number optional fields in the IOAM DEX option header. Two bits "N" and "I" (i.e., the third and fourth bits in the Extension-Flags field) are reserved to indicate the presence of the optional branch ID field. "N" stands for the Node ID and "I" stands for the interface index. If "N" and "I" are both set to 1, the optional multicast branch ID field is present; otherwise it is absent.

Θ	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	
+-+-+	-+-+-+-+-+-+-+-	+-+-+-+-	+-+-+-+-	+-+-+-+	-+-+-+-	
	Namespace-ID	I	Flags	F S N	I E-Flags	
+-+-+	-+-+-+-+-+-+-+-	+-+-+-+-	+-+-+-+-	+-+-+-+	-+-+-+-	
IOAM-Trace-Type				Res	erved	
+-						
Flow ID (optional)						
+-						
Sequence Number (Optional)						
+-						
Multicast Branch ID						
	(optional)					
+-						

Figure 3: Carry Branch ID in IOAM DEX option header

Once a node gets the branch ID information from the upstream, it MUST carry this information in its telemetry data export postcards, so the original multicast tree can be correctly reconstructed based on the postcards.

4.2. Per-section postcard for IOAM Trace

The second solution is a combination of the IOAM trace option and the postcard-based telemetry. To avoid data redundancy, at each branch fork node, the trace data accumulated up to this node is exported by a postcard before the packet is replicated. In this solution, each branch also needs to maintain some identifier to help correlate the postcards for each tree section. The natural way to accomplish this is to simply carry the branch fork node's data (including its ID) in the trace of each branch. This is also necessary because each replicated multicast packet can have different telemetry data pertaining to this particular copy (e.g., node delay, egress timestamp, and egress interface). As a consequence, the local data exported by each branch fork node can only contain partial data (e.g., ingress interface and ingress timestamp).

<u>Figure 4</u> shows an example in a segment of a multicast tree. Node B and D are two branch fork nodes and they will export a postcard covering the trace data for the previous section. The end node of each path will also need to export the data of the last section as a postcard.

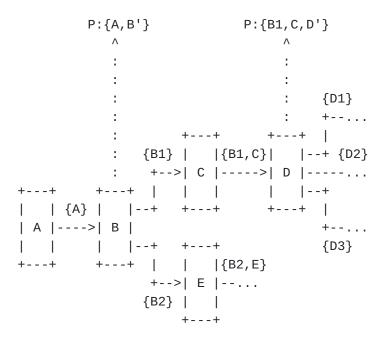


Figure 4: Per-section Postcard

There is no need to modify the IOAM trace option header format as specified in [RFC9197]. We just need to configure the branch fork nodes to export the postcards and refresh the IOAM header and data (e.g., clear the node data list and reset the Remaining Length field).

5. Application Considerations for Multicast Protocols

5.1. Mtrace verson 2

Mtrace version 2 (Mtrace2) [RFC8487] is a protocol that allows the tracing of an IP multicast routing path. Mtrace2 provides additional information such as the packet rates and losses, as well as other diagnostic information. Unlike unicast traceroute, Mtrace2 traces the path that the tree building messages follow from receiver to source. It is usually initiated from an Mtrace2 client by sending an Mtrace2 Query to a Last-Hop Router (LHR) or to a Rendezvous Point (RP). The LHR/RP turns the Query packet into an Mtrace2 Request, appends a Standard Response Block containing its interface addresses and packet statistics to the Request packet, and forwards the packet towards the source/RP. In a similar fashion, each router along the path to the source/RP appends a Standard Response Block to the end of the Request packet and forwards it to its upstream router. When the First-Hop Router (FHR) receives the Request packet, it appends its own Standard Response Block, turns the Request packet into a Reply, and unicasts the Reply back to the Mtrace2 client.

New on-path telemetry techniques will enhance Mtrace2, and other existing OAM solutions, with more granular and realtime network status data through direct measurements. There are various multicast protocols that are used to forward the multicast data. Each will require their own unique on-path telemetry solution. Mtrace2 doesn't integrate with IOAM directly, but network management systems may use Mtrace2 to learn about routers of interest.

5.2. Application in PIM

PIM-SM [RFC7761] is the most widely used multicast routing protocol deployed today. PIM-SSM, however, is the preferred method due to its simplicity and removal of network source discovery complexity. With PIM, control plane state is established in the network in order to forward multicast UDP data packets. PIM utilizes network based source discovery. PIM-SSM, however, utilizes application based source discovery. IP Multicast packets fall within the range of 224.0.0.0 through 239.255.255.255. The telemetry solution will need to work within this IPv4 address range and provide telemetry data for this UDP traffic.

A proposed solution for encapsulating the telemetry instruction header and metadata in IPv6 packets is described in [I-D.ietf-ippm-ioam-ipv6-options].

5.3. Application of MVPN X-PMSI Tunnel Encapsulation Attribute

IOAM, and the recommendations of this document, are equally applicable to multicast MPLS forwarded packets. Multipoint Label Distribution Protocol (mLDP), P2MP RSVP-TE, Ingress Replication (IR) and PIM MDT SAFI with GRE Transport are all commonly used within a Multicast VPN (MVPN) environment utilizing MVPN procedures such as Multicast in MPLS/BGP IP VPNs [RFC6513] and BGP Encoding and Procedures for Multicast in MPLS/BGP IP VPNs [RFC6514]. MLDP LDP Extension for P2MP and MP2MP LSPs [RFC6388] provides extensions to LDP to establish point-to-multipoint (P2MP) and multipoint-tomultipoint (MP2MP) label switched paths (LSPs) in MPLS networks. P2MP RSVP-TE provides extensions to RSVP-TE RSVP-TE for P2MP LSPs [RFC4875] to establish traffic-engineered P2MP LSPs in MPLS networks. Also used is <u>Ingress Replication Tunnels in Multicast VPNs</u> [RFC7988] which uses unicast replication from parent node to child node over a MPLS Unicast Tunnel. PIM MDT SAFI Multicast in BGP/MPLS IP VPNs [RFC6037]utilizes PIM-SSM, PIM-SM or PIM-BIDIR control plane with GRE transport data plane in the core for X-PMSI P-Tree using MVPN procedures. SR Replication Segment for Multi-point Service Delivery [I-D.ietf-spring-sr-replication-segment] utilizes replication segments for P2MP multicast service delivery in Segment Routing SR-MPLS networks. The telemetry solution will need to be able to follow these P2MP and MP2MP paths. The telemetry instruction header and data should be encapsulated into MPLS packets on P2MP and MP2MP paths. A corresponding proposal is described in [I-D.song-mpls-extension-header].

5.4. Application in BIER

<u>BIER</u> [<u>RFC8279</u>] adds a new header to multicast packets and allows the multicast packets to be forwarded according to the header only. By eliminating the requirement of maintaining per multicast group state, BIER is more scalable than the traditional multicast solutions.

<u>OAM Requirements for BIER [I-D.ietf-bier-oam-requirements</u>] lists many of the requirements for OAM at the BIER layer which will help in the forming of on-path telemetry requirements as well.

Depending on how the BIER header is encapsulated into packets with different transport protocols, the method to encapsulate the telemetry instruction header and metadata also varies. It is also possible to make the instruction header and metadata a part of the BIER header itself, such as in a TLV.

BIER-TE [RFC9262] contains multicast tree information in the packet header. It would therefore be possible to directly deduce the tree, that a packet traversed, when correlating received IOAM information.

6. Security Considerations

The schemes discussed in this document share the same security considerations for the IOAM trace option [RFC9197] and the IOAM DEX option [RFC9326]. In particular, since multicast has a built-in nature for packet amplification, the possible amplification risk for the DEX-based scheme is greater than the case of unicast. Hence, stricter mechanisms for protections need to be applied. In addition to selecting packets to enable DEX and limiting the exported traffic rate, we can also allows only a subset of the nodes in a multicast tree to process the option and export the data (e.g., only the branching nodes in the multicast tree are configured to process the option).

7. IANA Considerations

The document requests two new extension flag registrations in the "IOAM DEX Extension-Flags" registry, as described in Section 4.1.

Bit 2 "Multicast Branching Node ID [RFC XXXX] [RFC Editor: please replace with the RFC number of the current document]".

Bit 3 "Multicast Branching Interface Index [RFC XXXX] [RFC Editor: please replace with the RFC number of the current document]".

8. Acknowledgments

The authors would like to thank Frank Brockners, Nils Warnke, Jake Holland, Dino Farinacci, Henrik Nydell, and Toerless Eckert for their comments and suggestions.

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.
- [RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S.
 Yasukawa, Ed., "Extensions to Resource Reservation
 Protocol Traffic Engineering (RSVP-TE) for Point-to Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI
 10.17487/RFC4875, May 2007, https://www.rfc-editor.org/info/rfc4875.
- [RFC6388] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, DOI 10.17487/RFC6388, November 2011, https://www.rfc-editor.org/info/rfc6388>.
- [RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/ BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, https://www.rfc-editor.org/info/rfc6513>.
- [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>.
- [RFC7988] Rosen, E., Ed., Subramanian, K., and Z. Zhang, "Ingress Replication Tunnels in Multicast VPN", RFC 7988, DOI

- 10.17487/RFC7988, October 2016, <<u>https://www.rfc-</u>editor.org/info/rfc7988>.
- [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>.

- [RFC9326] Song, H., Gafni, B., Brockners, F., Bhandari, S., and T.
 Mizrahi, "In Situ Operations, Administration, and
 Maintenance (IOAM) Direct Exporting", RFC 9326, DOI
 10.17487/RFC9326, November 2022, https://www.rfc-editor.org/info/rfc9326.

9.2. Informative References

- [I-D.ietf-bier-oam-requirements] Mirsky, G., Nainar, N. K., Chen,
 M., and S. Pallagatti, "Operations, Administration and
 Maintenance (OAM) Requirements for Bit Index Explicit
 Replication (BIER) Layer", Work in Progress, Internet Draft, draft-ietf-bier-oam-requirements-13, 10 August
 2023, https://datatracker.ietf.org/doc/html/draft-ietf-bier-oam-requirements-13.

<https://datatracker.ietf.org/doc/html/draft-ietf-ippmioam-ipv6-options-12>.

[I-D.ietf-spring-sr-replication-segment]

Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. J. Zhang, "SR Replication segment for Multi-point Service Delivery", Work in Progress, Internet-Draft, draft-ietf-spring-sr-replication-segment-19, 28 August 2023, https://datatracker.ietf.org/doc/html/draft-ietf-spring-sr-replication-segment-19>.

[I-D.song-ippm-postcard-based-telemetry]

Song, H., Mirsky, G., Zhou, T., Li, Z., Graf, T., Mishra, G. S., Shin, J., and K. Lee, "On-Path Telemetry using Packet Marking to Trigger Dedicated OAM Packets", Work in Progress, Internet-Draft, draft-song-ippm-postcard-based-telemetry-16, 2 June 2023, https://datatracker.ietf.org/doc/html/draft-song-ippm-postcard-based-telemetry-16>.

- [I-D.song-mpls-extension-header] Song, H., Zhou, T., Andersson, L.,
 Zhang, Z. J., and R. Gandhi, "MPLS Network Actions using
 Post-Stack Extension Headers", Work in Progress,
 Internet-Draft, draft-song-mpls-extension-header-12, 14
 April 2023, https://datatracker.ietf.org/doc/html/draft-song-mpls-extension-header-12>.
- [RFC6037] Rosen, E., Ed., Cai, Y., Ed., and IJ. Wijnands, "Cisco Systems' Solution for Multicast in BGP/MPLS IP VPNs", RFC 6037, DOI 10.17487/RFC6037, October 2010, https://www.rfc-editor.org/info/rfc6037>.
- [RFC8487] Asaeda, H., Meyer, K., and W. Lee, Ed., "Mtrace Version
 2: Traceroute Facility for IP Multicast", RFC 8487, DOI
 10.17487/RFC8487, October 2018, https://www.rfc-editor.org/info/rfc8487>.

Authors' Addresses

Haoyu Song Futurewei Technologies 2330 Central Expressway Santa Clara, United States of America

Email: hsong@futurewei.com

Mike McBride Futurewei Technologies 2330 Central Expressway Santa Clara, United States of America

Email: mmcbride@futurewei.com

Greg Mirsky Ericsson

Email: gregimirsky@gmail.com

Gyan Mishra Verizon Inc.

Email: gyan.s.mishra@verizon.com

Hitoshi Asaeda National Institute of Information and Communications Technology Japan

Email: asaeda@nict.go.jp

Tianran Zhou Huawei Technologies Beijing China

Email: <u>zhoutianran@huawei.com</u>