

MBONED
Internet-Draft
Intended status: Standards Track
Expires: January 7, 2022

H. Song
M. McBride
Futurewei Technologies
G. Mirsky
ZTE Corp.
G. Mishra
Verizon Inc.
H. Asaeda
NICT
T. Zhou
Huawei
July 6, 2021

Multicast On-path Telemetry Solutions
draft-ietf-mboned-multicast-telemetry-01

Abstract

This document discusses the requirement of on-path telemetry for multicast traffic. The existing solutions are examined and their issues are identified. Solution modifications are proposed to allow the original multicast tree to be correctly reconstructed without unnecessary replication of telemetry information.

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 January 7, 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
2.	Requirements for Multicast Traffic Telemetry	3
3.	Issues of Existing Techniques	4
4.	Proposed Modifications to Existing Techniques	5
4.1.	Per-hop postcard using IOAM DEX	5
4.2.	Per-section postcard	7
5.	Considerations for Different Multicast Protocols	8
5.1.	Application in PIM	8
5.2.	Application of MVPN X-PMSI Tunnel Encapsulation Attribute	9
5.3.	Application in BIER	9
6.	Security Considerations	10
7.	IANA Considerations	10
8.	Contributors	10
9.	Acknowledgments	10
10.	References	10
10.1.	Normative References	10
10.2.	Informative References	12
	Authors' Addresses	13

[1.](#) Introduction

Multicast traffic is used across operator networks to support residential broadband customers, private MPLS customers and used with corporate intranet internal customers. 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 viewing experience. If a packet is dropped, and that packet happens to be a

reference frame (I-Frame) in the video feed, the client receiver of the multicast feed goes into buffering mode resulting in a frozen window. Multicast packet drops and delay can 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 [[I-D.ietf-ippm-ioam-data](#)], Postcard-based Telemetry [[I-D.song-ippm-postcard-based-telemetry](#)], and Hybrid Two-Step (HTS) [[I-D.mirsky-ippm-hybrid-two-step](#)] are useful and complementary to the existing active OAM performance monitoring methods, provide promising means to directly monitor the network experience of multicast traffic. However, multicast traffic has some unique characteristics which pose some challenges on efficiently applying such techniques.

The IP Multicast S,G data is identical from one branch to another on it's way to multiple receivers. When adding iOAM trace data, to multicast packets, we enlarge data packets thus consuming more network bandwidth. Instead of adding iOAM trace data, it could be more efficient to collect the telemetry information using solutions, such as iOAM postcard or HTS, to cut down on the redundant iOAM data. The problem is that a postcard type solution doesn't have a branch identifier.

This draft proposes a set of solutions to this iOAM data redundancy problem. 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 (MLDP, RSVP-TE) 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 node for efficiency.

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

- o Reconstruct and visualize the multicast tree through data plane monitoring.
- o Gather the multicast packet delay and jitter performance.

- o Find the multicast packet drop location and reason.
- o 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 and trace, may not be 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 to address the above mentioned requirements. The representative techniques include In-situ OAM (IOAM) Trace option [[I-D.ietf-ippm-ioam-data](#)], IOAM Direct Export (DEX) option [[I-D.ioamteam-ippm-ioam-direct-export](#)], and Postcard-based Telemetry with Packet Marking(PBT-M) [[I-D.song-ippm-postcard-based-telemetry](#)]. However, unlike unicast, multicast poses some unique challenges to applying these techniques.

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

If the IOAM trace option is used for on-path data collection, the partial trace data will also be replicated into multiple copies. The end result is that each copy of the multicast packet has a complete trace. Most of the data, however, is redundant. Data redundancy introduces unnecessary header overhead, wastes network bandwidth, and complicates the data processing. In case the multicast tree is large, and the path is long, the redundancy problem becomes severe.

The PBT solutions, including the IOAM DEX option and PBT-M, 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 the tree branches properly so it can bring confusion about the multicast tree topology. For example, Node A has two branches, one to Node B and the other to node D, and Node B leads to Node C and Node D leads to Node E. From the received postcards, one cannot tell whether or not Node C(E) is the next hop of Node B(D).

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. Proposed Modifications to Existing Techniques

Two solutions are proposed to address the above issues. One is built on PBT and requires augmentation or modification to the instruction header of the IOAM Direct Export Option; the other combines the IOAM trace option and PBT for an optimized solution.

4.1. Per-hop postcard using IOAM DEX

One way to mitigate PBT's multiple 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 uses an instruction header. To make the branch identifier globally unique, the branch node ID plus an index is used. For example, if Node A has two branches, one to Node B and one 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 unchanged for each multicast tree instance and carried with the multicast packet until the next branch node. Each postcard needs to include the branch identifier in the export data. The branch identifier, along with the other fields such as flow ID and sequence number, is sufficient for the data analyzer to reconstruct the topology of the multicast tree.

Figure 1 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.

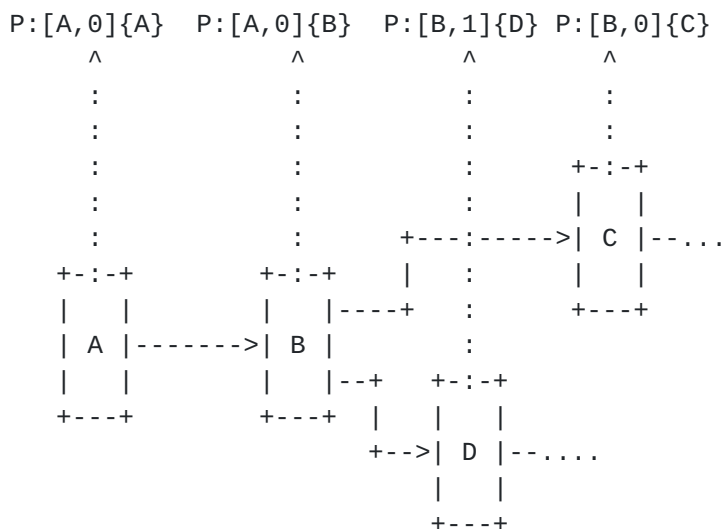


Figure 1: Per-hop Postcard

Each branch fork node needs to generate the branch ID for each branch in its multicast tree instance and include it in the IOAM DEX option header so the downstream node can learn it. The branch ID contains two parts: the branch fork node ID and a unique branch index.

Figure 2 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. A bit "M" in the Flags field is reserved to indicate the presence of the branch index field. The "M" flag position will be determined later after the other flags are specified in [\[I-D.ioamteam-ippm-ioam-direct-export\]](#).

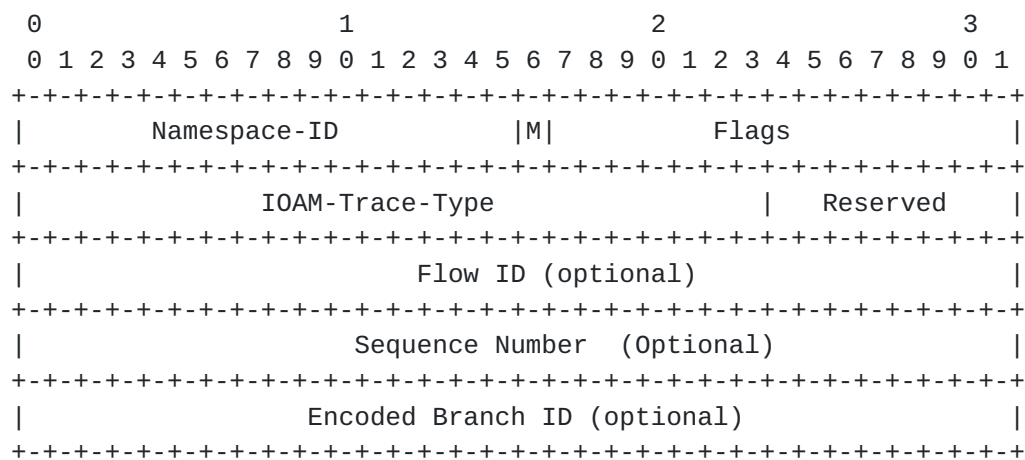


Figure 2: Carry Branch Index in IOAM DEX option header

To avoid introducing a new type of data field to the IOAM DEX option header, we can encode the branch identifier using the existing node ID data field as defined in [\[I-D.ietf-ippm-ioam-data\]](#). Currently, the node ID field occupies three octets. A simple solution is to shorten the node ID field so a number of bits can be saved to encode the branch index, as shown in Figure 3.

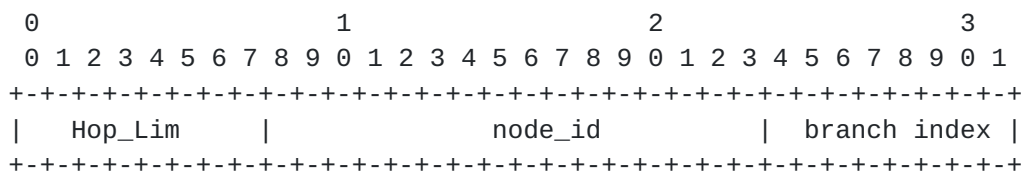


Figure 3: Encode Branch Index with Node ID Method 1

Another encoding method is to use the sum of the node ID and the branch index as the new node ID, as shown in Figure 4. As long as the node IDs are assigned with large enough gap, the telemetry data

analyzer can still successfully recover the original node ID and branch index.

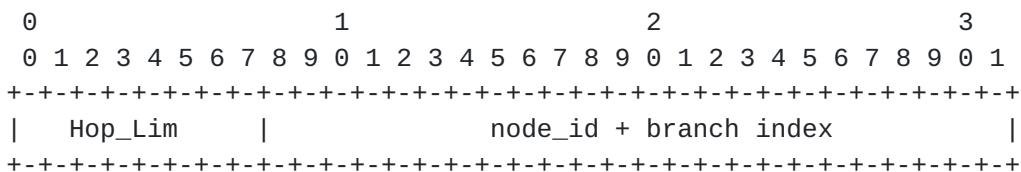


Figure 4: Encode Branch Index with Node ID Method 2

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

The second solution is a combination of the IOAM trace mode and PBT. To avoid data redundancy at each branch node, the trace data accumulated, to that point, is exported by a postcard before the packet is replicated. In this case, each branch still 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 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 node can only contain partial data (e.g., ingress interface and ingress timestamp).

Figure 5 shows an example in a segment of a multicast tree. Node B and D are two branch 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.

PIM-SM [RFC7761] is the most widely used multicast routing protocol deployed today. Of the various PIM modes (PIM-SM, PIM-DM, BIDIR-PIM, PIM-SSM), PIM-SSM is the preferred method due to its simplicity and removal of network source discovery complexity. With all PIM modes, control plane state is established in the network in order to forward multicast UDP data packets. All PIM modes utilize network based source discovery except for PIM-SSM, which 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 address range and provide telemetry data for this UDP traffic.

The proposed solutions for encapsulating the telemetry instruction header and metadata in IPv4/IPv6 UDP packets are described in [\[I-D.herbert-ipv4-udpencap-eh\]](#) and [\[I-D.ioametal-ippm-6man-ioam-ipv6-deployment\]](#).

5.2. Application of MVPN X-PMSI Tunnel Encapsulation Attribute

Multipoint Label Distribution Protocol (mLDP), P2MP RSVP-TE, Ingress Replication (IR), PIM MDT SAFI with GRE Transport, are commonly used within a Multicast VPN (MVPN) environment utilizing MVPN procedures 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-to-multipoint (MP2MP) label switched paths (LSPs) in MPLS networks. P2MP RSVP-TE provides extensions to RSVP-TE for P2MP LSPs [\[RFC4875\]](#) for establish traffic-engineered P2MP LSPs in MPLS networks. Ingress Replication (IR) P2MP Trees Ingress Replication Tunnels in Multicast VPNs [\[RFC7988\]](#) using unicast replication from parent node to child node over MPLS Unicast Tunnel. PIM MDT SAFI Multicast in BGP/MPLS IP VPNs [\[RFC6037\]](#) utilizes PIM modes PIM-SSM, PIM-SM, PIM-BIDIR control plane with GRE transport data plane in the core for X-PMSI P-Tree using MVPN procedures. Replication SID SR Replication Segment for Multi-point Service Delivery [\[I-D.ietf-spring-sr-replication-segment\]](#) 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 [\[I-D.song-mpls-extension-header\]](#).

5.3. Application in BIER

BIER [\[RFC8279\]](#) 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.

OAM Requirements for BIER [\[I-D.ietf-bier-oam-requirements\]](#) lists many of the requirements for OAM at the BIER layer which will help in the forming of on-path telemetry requirements as well.

There is also current work to provide solutions for BIER forwarding in ipv6 networks. For instance, a solution, BIER in Non-MPLS IPv6

Networks [[I-D.xie-bier-ipv6-encapsulation](#)], proposes a new bier Option Type codepoint from the "Destination Options and Hop-by-Hop Options" IPv6 sub-registry. This is similar to what IOAM proposes for IPv6 transport.

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.

6. Security Considerations

No new security issues are identified other than those discovered by the IOAM, PBT and HTS drafts.

7. IANA Considerations

The document makes no request of IANA.

8. Contributors

TBD

9. Acknowledgments

The authors would like to thank Frank Brockners for the comments and advice.

10. References

10.1. 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>>.
- [RFC4687] Yasukawa, S., Farrel, A., King, D., and T. Nadeau, "Operations and Management (OAM) Requirements for Point-to-Multipoint MPLS Networks", [RFC 4687](#), DOI 10.17487/RFC4687, September 2006, <<https://www.rfc-editor.org/info/rfc4687>>.

- [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>>.
- [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>>.
- [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>>.
- [RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", [RFC 6514](#), DOI 10.17487/RFC6514, February 2012, <<https://www.rfc-editor.org/info/rfc6514>>.
- [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, <<https://www.rfc-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>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", [RFC 8279](#), DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.

- [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>>.

10.2. Informative References

- [I-D.herbert-ipv4-udpencap-eh]
Herbert, T., "IPv4 Extension Headers and UDP Encapsulated Extension Headers", [draft-herbert-ipv4-udpencap-eh-01](#) (work in progress), March 2019.
- [I-D.ietf-bier-oam-requirements]
Mirsky, G., Kumar, N., Chen, M., and S. Pallagatti, "Operations, Administration and Maintenance (OAM) Requirements for Bit Index Explicit Replication (BIER) Layer", [draft-ietf-bier-oam-requirements-11](#) (work in progress), November 2020.
- [I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", [draft-ietf-ippm-ioam-data-12](#) (work in progress), February 2021.
- [I-D.ietf-spring-sr-replication-segment]
Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. Zhang, "SR Replication Segment for Multi-point Service Delivery", [draft-ietf-spring-sr-replication-segment-04](#) (work in progress), February 2021.
- [I-D.ioametal-ippm-6man-ioam-ipv6-deployment]
Bhandari, S., Brockners, F., Mizrahi, T., Kfir, A., Gafni, B., Spiegel, M., Krishnan, S., and M. Smith, "Deployment Considerations for In-situ OAM with IPv6 Options", [draft-ioametal-ippm-6man-ioam-ipv6-deployment-03](#) (work in progress), March 2020.
- [I-D.ioamteam-ippm-ioam-direct-export]
Song, H., Gafni, B., Zhou, T., Li, Z., Brockners, F., Bhandari, S., Sivakolundu, R., and T. Mizrahi, "In-situ OAM Direct Exporting", [draft-ioamteam-ippm-ioam-direct-export-00](#) (work in progress), October 2019.
- [I-D.mirsky-ippm-hybrid-two-step]
Mirsky, G., Lingqiang, W., Zhui, G., and H. Song, "Hybrid Two-Step Performance Measurement Method", [draft-mirsky-ippm-hybrid-two-step-09](#) (work in progress), March 2021.

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

Song, H., Mirsky, G., Filsfils, C., Abdelsalam, A., Zhou, T., Li, Z., Shin, J., and K. Lee, "Postcard-based On-Path Flow Data Telemetry using Packet Marking", [draft-song-ippm-postcard-based-telemetry-09](#) (work in progress), February 2021.

[I-D.song-mpls-extension-header]

Song, H., Li, Z., Zhou, T., and L. Andersson, "MPLS Extension Header", [draft-song-mpls-extension-header-04](#) (work in progress), April 2021.

[I-D.xie-bier-ipv6-encapsulation]

Xie, J., Geng, L., McBride, M., Asati, R., Dhanaraj, S., Zhu, Y., Qin, Z., Shin, M., Mishra, G., and X. Geng, "Encapsulation for BIER in Non-MPLS IPv6 Networks", [draft-xie-bier-ipv6-encapsulation-10](#) (work in progress), February 2021.

Authors' Addresses

Haoyu Song
Futurewei Technologies
2330 Central Expressway
Santa Clara
USA

Email: hsong@futurewei.com

Mike McBride
Futurewei Technologies
2330 Central Expressway
Santa Clara
USA

Email: mmcbride@futurewei.com

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com

Gyan Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

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

Email: asaeda@nict.go.jp

Tianran Zhou
Huawei

Email: zhoutianran@huawei.com

