

BIER
Internet-Draft
Intended status: Standards Track
Expires: October 29, 2018

Z. Zhang
A. Przygienda
Juniper Networks
A. Sajassi
Cisco Systems
J. Rabadan
Nokia
April 27, 2018

EVPN BUM Using BIER
draft-ietf-bier-evpn-01

Abstract

This document specifies protocols and procedures for forwarding broadcast, unknown unicast and multicast (BUM) traffic of Ethernet VPNs (EVPN) using Bit Index Explicit Replication (BIER).

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](#).

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 October 29, 2018.

Copyright Notice

Copyright (c) 2018 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.	Terminologies	3
2.	Use of the PMSI Tunnel Attribute	4
2.1.	Explicit Tracking	5
2.1.1.	Using IMET/SMET routes	5
2.1.2.	Using S-PMSI/Leaf A-D Routes	6
2.2.	MPLS Label in PTA	7
3.	Multihoming Split Horizon	7
4.	Data Plane	8
4.1.	Encapsulation and Transmission	8
4.1.1.	At a BFIR that is an Ingress PE	8
4.1.2.	At a BFIR that is a P-tunnel Segmentation Point	10
4.2.	Disposition	10
4.2.1.	At a BFER that is an Egress PE	10
4.2.2.	At a BFER that is a P-tunnel Segmentation Point	11
5.	IANA Considerations	11
6.	Security Considerations	11
7.	Acknowledgements	11
8.	References	11
8.1.	Normative References	11
8.2.	Informative References	12
	Authors' Addresses	13

1. Introduction

[RFC7432] and [[RFC8365](#)] specify the protocols and procedures for Ethernet VPNs (EVPNs). For broadcast, unknown unicast and multicast (BUM) traffic, provider/underlay tunnels (referred to as P-tunnels) are used to carry the BUM traffic. Several kinds of tunnel technologies can be used, as specified in [[RFC7432](#)].

Bit Index Explicit Replication (BIER) ([[RFC8279](#)]) is an architecture that provides optimal multicast forwarding through a "multicast domain", without requiring intermediate routers to maintain any per-flow state or to engage in an explicit tree-building protocol. The

purpose of this document is to specify the protocols and procedures to transport EVPN BUM traffic using BIER.

The EVPN BUM procedures specified in [[RFC7432](#)] and extended in [[I-D.ietf-bess-evpn-bum-procedure-updates](#)], [[I-D.ietf-bess-evpn-igmp-ml-d-proxy](#)], and [[I-D.zhang-bess-mvpn-evpn-cmcast-enhancements](#)] are much aligned with MVPN procedures. As such, this document is also very much aligned with [[I-D.ietf-bier-mvpn](#)]. For terseness, some background, terms and concepts are not repeated here. Additionally, some text is borrowed verbatim from [[I-D.ietf-bier-mvpn](#)].

1.1. Terminologies

- o BFR: Bit-Forwarding Router.
- o BFIR: Bit-Forwarding Ingress Router.
- o BFER: Bit-Forwarding Egress Router.
- o BFR-Prefix: An IP address that uniquely identifies a BFR and is routeable in a BIER domain.
- o C-S: A multicast source address, identifying a multicast source located at a VPN customer site.
- o C-G: A multicast group address used by a VPN customer.
- o C-flow: A customer multicast flow. Each C-flow is identified by the ordered pair (source address, group address), where each address is in the customer's address space. The identifier of a particular C-flow is usually written as (C-S,C-G). Sets of C-flows can be identified by the use of the "C-*" wildcard (see [[RFC6625](#)]), e.g., (C-*,C-G).
- o P-tunnel. A multicast tunnel through the network of one or more SPs. P-tunnels are used to transport MVPN multicast data
- o IMET Route: Inclusive Multicast Ethernet Tag Auto-Discovery route. Carried in BGP Update messages, these routes are used to advertise the "default" P-tunnel for a particular broadcast domain.
- o SMET Route: Selective Multicast Ethernet Tag Auto-Discovery route. Carried in BGP Update messages, these routes are used to advertise the C-flows that the advertising PE is interested in.
- o S-PMSI A-D route: Selective Provider Multicast Service Interface Auto-Discovery route. Carried in BGP Update messages, these

routes are used to advertise the fact that particular C-flows are bound to (i.e., are traveling through) particular P-tunnels.

- o PMSI Tunnel attribute (PTA). This BGP attribute carried is used to identify a particular P-tunnel. When C-flows of multiple VPNs are carried in a single P-tunnel, this attribute also carries the information needed to multiplex and demultiplex the C-flows.

2. Use of the PMSI Tunnel Attribute

[RFC7432] specifies that Inclusive Multicast Ethernet Tag (IMET) routes carry a PMSI Tunnel Attribute (PTA) to identify the particular P-tunnel to which one or more BUM flows are being assigned, the same as specified in [RFC6514] for MVPN. [I-D.ietf-bier-mvpn] specifies the encoding of PTA for use of BIER with MVPN. Much of that specification is reused for use of BIER with EVPN and much of the text below is borrowed verbatim from [I-D.ietf-bier-mvpn].

The PMSI Tunnel Attribute (PTA) contains the following fields:

- o "Tunnel Type". The same codepoint 0x0B that IANA has assigned for [I-D.ietf-bier-mvpn] for the new tunnel type "BIER" is used for EVPN as well.
- o "Tunnel Identifier". When the "tunnel type" field is "BIER", this field contains two subfields. The text below is exactly as in [I-D.ietf-bier-mvpn].
 - 1 The first subfield is a single octet, containing the sub-domain-id of the sub-domain to which the BFIR will assign the packets that it transmits on the PMSI identified by the NLRI of the IMET, S-PMSI A-D, or per-region I-PMSI A-D route that contains this PTA. How that sub-domain is chosen is outside the scope of this document.
 - 2 The second subfield is a two-octet field containing the BFR-id, in the sub-domain identified in the first subfield, of the router that is constructing the PTA.
 - 3 The third subfield is the BFR-Prefix (see [RFC8279]) of the originator of the route that is carrying this PTA. This will either be a /32 IPv4 address or a /128 IPv6 address. Whether the address is IPv4 or IPv6 can be inferred from the total length of the PMSI Tunnel attribute.

The BFR-prefix need not be the same IP address that is carried in any other field of the x-PMSI A-D route, even if the BFIR is the originating router of the x-PMSI A-D route.

- o "MPLS label". For EVPN-MPLS [[RFC7432](#)], this field contains an upstream-assigned MPLS label. It is assigned by the BFIR. Constraints on the way in which the originating router selects this label are discussed in [Section 2.2](#). For EVPN-VXLAN/NVGRE [[RFC8365](#)], this field is a 24-bit VNI/VSID of global significance.
- o "Flags". When the tunnel type is BIER, two of the flags in the PTA Flags field are meaningful. Details about the use of these flags can be found in [Section 2.1](#).
 - * "Leaf Info Required per Flow (LIR-pF)"
[[I-D.ietf-bess-mvpn-expl-track](#)]
 - * "Leaf Info Required Bit (LIR)"

Note that if a PTA specifying "BIER" is attached to an IMET, S-PMSI A-D, or per-region I-PMSI A-D route, the route MUST NOT be distributed beyond the boundaries of a BIER domain. That is, any routers that receive the route must be in the same BIER domain as the originator of the route. If the originator is in more than one BIER domain, the route must be distributed only within the BIER domain in which the BFR-Prefix in the PTA uniquely identifies the originator. As with all MVPN routes, distribution of these routes is controlled by the provisioning of Route Targets.

[2.1. Explicit Tracking](#)

When using BIER to transport an EVPN BUM data packet through a BIER domain, an ingress PE functions as a BFIR (see [[RFC8279](#)]). The BFIR must determine the set of BFERs to which the packet needs to be delivered. This can be done in either of two ways in the following two sections.

[2.1.1. Using IMET/SMET routes](#)

Both IMET and SMET (Selective Multicast Ethernet Tag [[I-D.ietf-bess-evpn-igmp-mld-proxy](#)]) routes provide explicit tracking functionality.

For an inclusive PMSI, the set of BFERs to deliver traffic to includes the originators of all IMET routes for a broadcast domain. For a selective PMSI, the set of BFERs to deliver traffic to includes the originators of corresponding SMET routes.

The SMET routes do not carry a PTA. When an ingress PE sends traffic on a selective tunnel using BIER, it uses the upstream assigned label that is advertised in its IMET route.

Only when selectively forwarding is for all flows without tunnel segmentation, SMET routes are used without the need for S-PMSI A-D routes. Otherwise, the procedures in the following section apply.

2.1.2. Using S-PMSI/Leaf A-D Routes

There are two cases where S-PMSI/Leaf A-D routes are used as discussed in the following two sections.

2.1.2.1. Selective Forwarding Only for Some Flows

With the SMET procedure, a PE advertises an SMET route for each (C-S,C-G) or (C-*,C-G) state that it learns on its ACs, and each SMET route is tracked by every PE in the same broadcast domain. It may be desired that SMET routes are not used to reduce the burden of explicit tracking.

In this case, most multicast traffic will follow the I-PMSI (advertised via IMET route) and only some flows follow S-PMSIs. To achieve that, S-PMSI/Leaf A-D routes can be used, as specified in [[I-D.ietf-bess-evpn-bum-procedure-updates](#)].

The rules specified in [Section 2.2.1](#) and Section 2.2.2 of [[I-D.ietf-bier-mvpn](#)] apply.

2.1.2.2. Tunnel Segmentation

Another case where S-PMSI/Leaf A-D routes are necessary is tunnel segmentation, which is also specified in [[I-D.ietf-bess-evpn-bum-procedure-updates](#)], and further clarified in [[I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements](#)] for segmentation with SMET routes. This is only applicable to EVPN-MPLS.

The rules specified in Section 2.2.1 of [[I-D.ietf-bier-mvpn](#)] apply. Section 2.2.2 of [[I-D.ietf-bier-mvpn](#)] do not apply, because similar to MVPN, the LIR-pF flag cannot be used with segmentation.

2.1.2.3. Applicability of Additional MVPN Specifications

As with the MVPN case, Section "3. Use of the PMSI Tunnel Attribute in Leaf A-D routes" of [[I-D.ietf-bier-mvpn](#)] apply.

Notice that, [[I-D.ietf-bier-mvpn](#)] refers to procedures specified in [[RFC6625](#)] and [[I-D.ietf-bess-mvpn-expl-track](#)]. Those two documents were specified for MVPN but are actually applicable to IP multicast payload in EVPN as well.

2.2. MPLS Label in PTA

Rules in section 2.1 of [[I-D.ietf-bier-mvpn](#)] apply, EXCEPT the following three bullets (they do NOT apply to EVPN) in that section:

- o If the two routes do not have the same Address Family Identifier (AFI) value, then their respective PTAs MUST contain different MPLS label values. This ensures that when an egress PE receives a data packet with the given label, the egress PE can infer from the label whether the payload is an IPv4 packet or an IPv6 packet.
- o If the BFIR is an ingress PE supporting MVPN extranet ([\[RFC7900\]](#)) functionality, and if the two routes originate from different VRFs on this ingress PE, then the respective PTAs of the two routes MUST contain different MPLS label values.
- o If the BFIR is an ingress PE supporting the "Extranet Separation" feature of MVPN extranet (see [Section 7.3 of \[RFC7900\]](#)), and if one of the routes carries the "Extranet Separation" extended community but the other does not, then the respective PTAs of the two routes MUST contain different MPLS label values.

3. Multihoming Split Horizon

For EVPN-MPLS, [\[RFC7432\]](#) specifies the use of ESI labels to identify the ES from which a BUM packet originates. A PE receiving that packet from the core side will not forward it to the same ES. The procedure works for both Ingress Replication (IR) and RSVP-TE/mLDP P2MP tunnels, using downstream- and upstream-assigned ESI labels respectively. For EVPN-VXLAN/NVGRE, [\[RFC8365\]](#) specifies local-bias procedures, with which a PE receiving a BUM packet from the core side knows from encapsulation the ingress PE so it does not forward the packet to any multihoming ESes that the ingress PE is on, because the ingress PE already forwarded the packet to those ESes, regardless of whether the ingress PE is a DF for those ESes.

With BIER, the local-bias procedure still applies for EVPN-VXLAN/NVGRE as the BFIR-id in the BIER header identifies the ingress PE. For EVPN-MPLS, ESI label procedures also still apply though two upstream assigned labels will be used (one for identifying the broadcast domain and one for identifying the ES) - the same as in the case of using a single P2MP tunnel for multiple broadcast domains. The BFIR-id in the BIER header identifies the ingress PE that assigned those two labels.

4. Data Plane

Similar to MVPN, the EVPN application plays the role of the "multicast flow overlay" as described in [[RFC8279](#)].

4.1. Encapsulation and Transmission

A BFIR could be either an ingress PE or a P-tunnel segmentation point. The procedures are slightly different as described below.

4.1.1. At a BFIR that is an Ingress PE

To transmit a BUM data packet, an ingress PE first determines the route matched for transmission and routes for tracking leaves according to the following rules.

1. If selective forwarding is not used, or it is not an IP Multicast packet after the ethernet header, the IMET route originated for the BD by the ingress PE is the route matched for transmission. Leaf tracking routes are all other received IMET routes for the BD.
2. Otherwise, if selective forwarding is used for all IP Multicast traffic based on SMET routes, the IMET route originated for the BD by the ingress PE is the route matched for transmission. Received SMET routes for the BD that best match the source and destination IP address are leaf tracking routes.
3. Otherwise, route matched for transmission is the S-PMSI A-D route originated by the ingress PE for the BD, that best matches the packet's source and destination IP address and has a PTA specifying a valid tunnel type that is not "no tunnel info". Leaf tracking routes are determined as following:
 - 1) If the match for transmission route carries a PTA that has the LIR flag set but does not have the LIR-pF flag set, the routes matched for tracking are Leaf A-D routes whose "route key" field is identical to the NLRI of the S-PMSI A-D route.
 - 2) If the match for transmission route carries a PTA that has the LIR-pF flag, the leaf tracking routes are Leaf A-D routes whose "route key" field is derived from the NLRI of the S-PMSI A-D route according to the procedures described in Section 5.2 of [[I-D.ietf-bess-mvpn-expl-track](#)].

Note that in both cases, SMET routes may be used in lieu of Leaf A-D routes, as a PE may omit the Leaf A-D route in response to an S-PMSI A-D route with LIR or LIR-pF bit set, if an SMET route

with the corresponding Tag, Source and Group fields is already originated [[I-D.ietf-bess-evpn-bum-procedure-updates](#)]. In particular, in the second case above, even though the SMET route does not have a PTA attached, it is still considered as a Leaf A-D route in response to a wildcard S-PMSI A-D route with the LIR-pF bit set.

4. Otherwise, route matched for transmission and leaf tracking routes are determined as in rule 1.

If no route is matched for transmission, the packet is not forwarded onto a p-tunnel. If the tunnel that the ingress determines to use based on the route matched for transmission (and considering interworking with PEs that do not support certain tunnel types per procedures in [[I-D.ietf-bess-evpn-igmp-mld-proxy](#)]) requires leaf tracking (e.g. Ingress Replication, RSVP-TE P2MP tunnel, or BIER) but there are no leaf tracking routes, the packet will not be forwarded onto a p-tunnel either.

The following text assumes that BIER is the determined tunnel type. The ingress PE pushes an upstream assigned ESI label per [[RFC7432](#)] if the following conditions are all met:

- o The packet is received on a multihomed ES.
- o It's EVPN-MPLS.
- o ESI label procedure is used for split-horizon.

The (upstream-assigned) MPLS label from the PTA of the route matched for transmission is then pushed onto the packet's label stack in case of EVPN-MPLS. In case of EVPN-VXLAN/NVGRE, a VXLAN/NVGRE header is prepended to the packet with the VNI/VSID set to the value in the PTA's label field and no IP/UDP header is used.

Then the packet is encapsulated in a BIER header and forwarded, according to the procedures of [[RFC8279](#)] and [[RFC8296](#)]. See especially [Section 4](#), "Imposing and Processing the BIER Encapsulation", of [[RFC8296](#)]. The "Proto" field in the BIER header is set to 2 in case of EVPN-MPLS or a value to be assigned in case of EVPN-VXLAN/NVGRE ([Section 5](#)).

In order to create the proper BIER header for a given packet, the BFIR must know all the BFERs that need to receive that packet. This is determined from the set of leaf tracking routes.

4.1.2. At a BFIR that is a P-tunnel Segmentation Point

In this case, the encapsulation for upstream segment of the p-tunnel includes (among other things) a label that identifies the x-PMSI or IMET A-D route that is the match for reception on the upstream segment. The segmentation point re-advertised the route into one or more downstream regions. Each instance of the re-advertised route for a downstream region has a PTA that specify tunnel information that is the same as or different from that of the route for a different region. For any particular downstream region, the route matched for transmission is the re-advertised route, and the leaf tracking routes are determined as following if needed for the tunnel type:

- o If the route matched for transmission is an x-PMSI route, it must have the LIR flag set in its PTA and the leaf tracking routes are all the matching Leaf A-D and SMET routes received in the downstream region.
- o If the route matched for transmission is an IMET route, the leaf tracking routes are all the IMET routes for the same BD received in the downstream region.

If the downstream region uses BIER, the packet is forwarded as following: the upstream segmentation's encapsulation is removed and the above mentioned label is swapped to the upstream-assigned label in the PTA of the route matched for transmission, and then a BIER header is imposed as in [Section 4.1.1](#).

4.2. Disposition

The same procedures in section 4.2 of [[I-D.ietf-bier-mvpn](#)] are followed for EVPN-MPLS, except some EVPN specifics discussed in the following two sub-sections in this document.

For EVPN-VXLAN/NVGRE, the only difference is that the payload is VXLAN/NVGRE and the VNI/VSID field in the VXLAN/NVGRE header is used to determine the corresponding mac VRF or broadcast domain.

4.2.1. At a BFER that is an Egress PE

Once the corresponding mac VRF or broadcast domain is determined from the upstream assigned label or VNI/VSID, EVPN forwarding procedures per [[RFC7432](#)] or [[RFC8365](#)] are followed. In case of EVPN-MPLS, if there is an inner label in the label stack following the BIER header, that inner label is considered as the upstream assigned ESI label for split horizon purpose.

4.2.2. At a BFER that is a P-tunnel Segmentation Point

This is only applicable to EVPN-MPLS. The same procedures in Section 4.2.2 of [[I-D.ietf-bier-mvpn](#)] are followed, subject to multihoming procedures specified in [[I-D.ietf-bess-evpn-bum-procedure-updates](#)].

5. IANA Considerations

This document requests two assignments in "BIER Next Protocol Identifiers" registry, with the following two recommended values:

- o 7: Payload is VXLAN encapsulated (no IP/UDP header)
- o 8: Payload is NVGRE encapsulated (no IP header)

6. Security Considerations

To be updated.

7. Acknowledgements

The authors thank Eric Rosen for his review and suggestions. Additionally, much of the text is borrowed verbatim from [[I-D.ietf-bier-mvpn](#)].

8. References

8.1. Normative References

- [[I-D.ietf-bess-evpn-bum-procedure-updates](#)]
Zhang, Z., Lin, W., Rabadan, J., Patel, K., and A. Sajassi, "Updates on EVPN BUM Procedures", [draft-ietf-bess-evpn-bum-procedure-updates-03](#) (work in progress), April 2018.
- [[I-D.ietf-bess-evpn-igmp-mld-proxy](#)]
Sajassi, A., Thoria, S., Patel, K., Yeung, D., Drake, J., and W. Lin, "IGMP and MLD Proxy for EVPN", [draft-ietf-bess-evpn-igmp-mld-proxy-01](#) (work in progress), March 2018.
- [[I-D.ietf-bess-mvpn-expl-track](#)]
Dolganow, A., Kotalwar, J., Rosen, E., and Z. Zhang, "Explicit Tracking with Wild Card Routes in Multicast VPN", [draft-ietf-bess-mvpn-expl-track-09](#) (work in progress), April 2018.

[I-D.ietf-bier-mvpn]

Rosen, E., Sivakumar, M., Aldrin, S., Dolganow, A., and T. Przygienda, "Multicast VPN Using BIER", [draft-ietf-bier-mvpn-11](#) (work in progress), March 2018.

[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>>.

[RFC6625] Rosen, E., Ed., Rekhter, Y., Ed., Hendrickx, W., and R. Qiu, "Wildcards in Multicast VPN Auto-Discovery Routes", [RFC 6625](#), DOI 10.17487/RFC6625, May 2012, <<https://www.rfc-editor.org/info/rfc6625>>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", [RFC 7432](#), DOI 10.17487/RFC7432, February 2015, <<https://www.rfc-editor.org/info/rfc7432>>.

[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>>.

[RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", [RFC 8296](#), DOI 10.17487/RFC8296, January 2018, <<https://www.rfc-editor.org/info/rfc8296>>.

8.2. Informative References

[I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements]

Zhang, Z., Kebler, R., Lin, W., and E. Rosen, "MVPN/EVPN C-Multicast Routes Enhancements", [draft-zzhang-bess-mvpn-evpn-cmcast-enhancements-00](#) (work in progress), July 2016.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", [RFC 8365](#), DOI 10.17487/RFC8365, March 2018, <<https://www.rfc-editor.org/info/rfc8365>>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks

EMail: zzhang@juniper.net

Antoni Przygienda
Juniper Networks

EMail: prz@juniper.net

Ali Sajassi
Cisco Systems

EMail: sajassi@cisco.com

Jorge Rabadan
Nokia

EMail: jorge.rabadan@nokia.com

