

BIER
Internet-Draft
Intended status: Standards Track
Expires: June 10, 2022

Z. Zhang
A. Przygienda
Juniper Networks
A. Sajassi
Cisco Systems
J. Rabadan
Nokia
December 7, 2021

EVPN BUM Using BIER
draft-ietf-bier-evpn-05

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 June 10, 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

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 [\[I-D.ietf-bess-evpn-bum-procedure-updates\]](#), [\[I-D.ietf-bess-evpn-igmp-mld-proxy\]](#), and [\[I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements\]](#) are much aligned with MVPN procedures. As such, this document is also very much aligned with [\[RFC8556\]](#). For terseness, some background, terms and concepts are not repeated here. Additionally, some text is borrowed verbatim from [\[RFC8556\]](#).

[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. [\[RFC8556\]](#) 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 [\[RFC8556\]](#).

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

- o "Tunnel Type". The same codepoint 0x0B that IANA has assigned for [\[RFC8556\]](#) 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 [\[RFC8556\]](#).

- 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 2.3](#). For EVPN-VXLAN/NVGRE/GENEVE [[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 2.2](#).
 - * "Leaf Info Required per Flow (LIR-pF)" [[RFC8534](#)]
 - * "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.](#) IP Based Tunnel and BIER PHP

When VXLAN/NVGRE/GENEVE is used for EVPN, by default the outer IP header (and UDP header in case of VXLAN/GENVE) is not included in the BIER payload, except when it is known apriori that BIER PHP [I-D.ietf-bier-php] is used in the BIER domain and the encapsulation (after BIER header is popped) between the BIER Penultimate Hop and the egress PE does not have a way to indicate the next header is VXLAN/NVGRE/GENEVE. In that case the full VXLAN/NVGRE/GENEVE encapsulation with an IP header MUST be included in the BIER payload. A well-known IP multicast address (to be assigned by IANA) is used as the destination address and the egress PEs MUST be set up to receive and process packets addressed to the address. The address is used for all BDs, and the inner VXLAN/NVGRE/GENEVE header will be used to identify BDs.

[2.2.](#) 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.

[2.2.1.](#) 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

P2MP tunnels, using downstream- and upstream-assigned ESI labels respectively. For EVPN-VXLAN/NVGRE/GENEVE, [\[RFC8365\]](#) 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/GENEVE 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 \[RFC8534\]](#).

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.

Internet-Draft

bier-evpn

December 2021

The MPLS label from the PTA of the route matched for transmission is then pushed onto the packet's label stack for EVPN-MPLS. For EVPN-VXLAN/NVGRE/GENEVE, a VXLAN/NVGRE/GENEVE header is prepended to the packet with the VNI/VSID set to the value in the PTA's label field, and then an IP/UDP header is prepended if needed (e.g. for PHP purpose).

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/GENEVE ([Section 5](#)) when IP header is not used, or 4/6 if IP header is used for EVPN-VXLAN/NVGRE/GENEVE.

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

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