

BIER Workgroup  
Internet Draft  
Intended status: Standard Track

H. Bidgoli, Ed.  
A. Dolganow  
Nokia  
Fengman Xu  
Verizon

Expires: December 23, 2017

June 21, 2017

PIM Tunneling Through BIER Core  
draft-hfa-bier-pim-tunneling-00

Abstract

Bit Index Explicit Replication (BIER) is an architecture that provides multicast forwarding through a "BIER domain" without requiring intermediate routers to maintain multicast related per-flow state. Neither does BIER require an explicit tree-building protocol for its operation. A multicast data packet enters a BIER domain at a "Bit-Forwarding Ingress Router" (BFIR), and leaves the BIER domain at one or more "Bit-Forwarding Egress Routers" (BFERs). The BFIR router adds a BIER header to the packet. Such header contains a bit-string in which each bit represents exactly one BFER to forward the packet to. The set of BFERs to which the multicast packet needs to be forwarded is expressed by the according set of bits switched on in BIER packet header.

This document describes the procedure needed for PIM to be tunneled through a BIER core. Allowing access CEs or PEs to run traditional PIM multicast services including draft rosen multicast MVPN through a core of BIER.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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." The list

of current Internet-Drafts can be accessed at  
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on October 8, 2017.

#### Copyright Notice

Copyright (c) 2017 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 (<http://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. Conventions used in this document . . . . .	4
2.1. Definitions . . . . .	4
3. PIM Tunneling Through BIER domain . . . . .	5
3.1. PIM-BFIR procedure . . . . .	6
3.1.1. BIER packet construction at PIM BFIR . . . . .	6
3.2. Tunneling PIM through the BIER domain procedure . . . . .	7
3.3. PIM-BFER procedure . . . . .	7
4. Datapath Forwarding . . . . .	8
4.1. BIER reverse path forwarding table . . . . .	8
4.2. Datapath traffic flow . . . . .	9
5. PIM-ASM behavior . . . . .	9
6. Draft Rosen multicast vpn behavior . . . . .	10
7. Security Considerations . . . . .	10
7.1. Normative References . . . . .	10
7.2. Informative References . . . . .	10
8. Acknowledgments . . . . .	11
Authors' Addresses . . . . .	11

#### 1. Introduction

Most Service Providers understand the benefit of BIER and would like a Core network that supports scalable multicast solution by removing the multicast states and deploying BIER.

That said greenfield deployment of BIER might not be possible for providers that deploy MVPN technology, or have more than 256 PEs in their network. Consider the following:

1. Most service providers deploy MVPN technology for multicast today. Their network structure typically is a Core network, which connects edge networks containing PEs. These PEs run MVPN services like draft rosen multicast vpns. In a typical tier one provider the number of PEs is well beyond 1K of routers. As the edge network expands the scale of multicast states in the core could test the routers limits. As such it is attractive to create a stateless BIER core which can transport MVPN technology. By deploying BIER in the core of the network the bottle necks for multicast states is removed. By pushing the multicast states to the edge provider (P) routers, a more manageable multicast state can be achieved.
2. Deploying greenfield BIER services for most providers could be a challenge. It might be attractive to deploy bier in multiple phases. Starting from the core of the network to remove the massive multicast state generated by traditional MVPN services is an ideal evolution path. By tunneling traditional MVPN technology through a BIER core an scalable and manageable network can be created.
3. Most vendors support 256 bits in the BIER header. Identifying only 256 PEs or CEs via a single BIER packet. Scaling beyond 256 PEs or CEs "might" require packet duplication depending on network topology. This packet duplication will be done via BIER Set Index (SI) which is explained in draft-ietf-bier-architecture. In the cases that packet duplication can't be avoid it might be desirable to segment the network to traditional MVPN technology at the access and BIER in the core. By moving the Bier in the core all core routers could be presented via the 256 bits in the BIER header.

In all above cases it might be attractive to be able to tunnel traditional MVPN services over a BIER core.

This draft explains the procedure to tunnel PIM through a BIER core, as such enable tunneling of traditional MVPN services like draft-rosen multicast vpns through a core of BIER.

The procedures of PIM tunneling should be used at the BIER edge routers. The BIER edge routers (BER) are connected to legacy PIM routers on one side and BIER routers on the other side. PIM routers continue to send PIM state messages to the BER but the BER does not

propagate PIM packets natively into the BIER sub-domain. Instead it will tunnel the PIM through BIER network.

In this draft the BFIR and BFER are the BER from multicast traffic point of view and not PIM signaling. That said the PIM BFIR (P-BFIR) and PIM (P-BFER) are BER from PIM signaling point of view.

As such a P-BFIR would be a BFER of the multicast traffic and a P-BFER would be a BFIR of the multicast traffic.

## 2. Conventions used in this document

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 RFC 2119 [RFC2119].

### 2.1. Definitions

Some of the terminology specified in [I-D.draft-ietf-bier-architecture-05] is replicated here and extended by necessary definitions:

#### BIER:

Bit Index Explicit Replication (The overall architecture of forwarding multicast using a Bit Position).

#### BFR:

Bit Forwarding Router (A router that participates in Bit Index Multipoint Forwarding). A BFR is identified by a unique BFR-prefix in a BIER domain.

#### BFIR:

Bit Forwarding Ingress Router (The ingress border router that inserts the BM into the packet). Each BFIR must have a valid BFR-id assigned. In this draft BIER will be used for forwarding and tunneling of control plain packet (i.e. PIM) and forwarding dataplain packets. BFIR is term used for dataplain packet forwarding.

#### BFER:

Bit Forwarding Egress Router. A router that participates in Bit Index Forwarding as leaf. Each BFER must be a BFR. Each BFER must have a valid BFR-id assigned. In this draft BIER will be used for forwarding and tunneling of control plain

packet (i.e. PIM) and forwarding dataplain packets. BFIR is term used for dataplain packet forwarding.

P-BFIR:

PIM-Bit Forwarding ingress router. Ingress boundary router between PIM domain and BIER domain. It tunnels PIM packet through a BIER domain toward the source.

P-BFER:

PIM-Bit Forwarding egress router. Egress boundary router between BIER domain and PIM domain. It decapsulates PIM packet from a BIER tunnel and forwards it to the PIM domain.

BRT:

BIER RPF Table, is built on the P-BFER. It tracks which P-BFIR is interested in a group. It is used to map the group to the P-BFIR BIER prefix.

BFT:

Bit Forwarding Tree used to reach all BFERs in a domain.

BIFT:

Bit Index Forwarding Table.

BIER sub-domain:

A further distinction within a BIER domain identified by its unique sub-domain identifier. A BIER sub-domain can support multiple BitString Lengths.

BFR-id:

An optional, unique identifier for a BFR within a BIER sub-domain.

### 3. PIM Tunneling Through BIER domain

Suppose BIER sub-domain is to be an IGP area or instance. The BIER edge routers (BER) can be ABRs that are connected to edge network via IGP or BGP, or they can be any provider (P) router that is selected to act as the BER in that BIER sub-domain.

Each BER is configured as per BIER requirements explained in draft-ietf-bier-architecture.

The BERs receive PIM joins from the downstream routers because they are on the path toward the source. Additionally, on these BERs all interfaces which are PIM enabled are configured to tunnel PIM over BIER.

### 3.1. PIM-BFIR procedure

When PIM joins for a certain (S,G) arrives on a BER, in this case the P-BFIR (BFIR from PIM signaling point of view). This router first find the route to the source. The route to the source is assumed to be an IGP route. The BER tries to resolve the source (S), in the process it of resolving the source the SPF calculation can return the P-BFER that is in the path to the source.

The procedure to find the P-BFER (BFER from PIM signaling point of view) can be via 2 mechanism and is beyond the scope of this draft.

- 1.The P-BFER can be an ABR or ASBR router which is summarizing the route to the source, and as such is the source of this route.
- 2.The P-BFER can be resolved via SPF calculation and finding the first BFIR in the path the source.

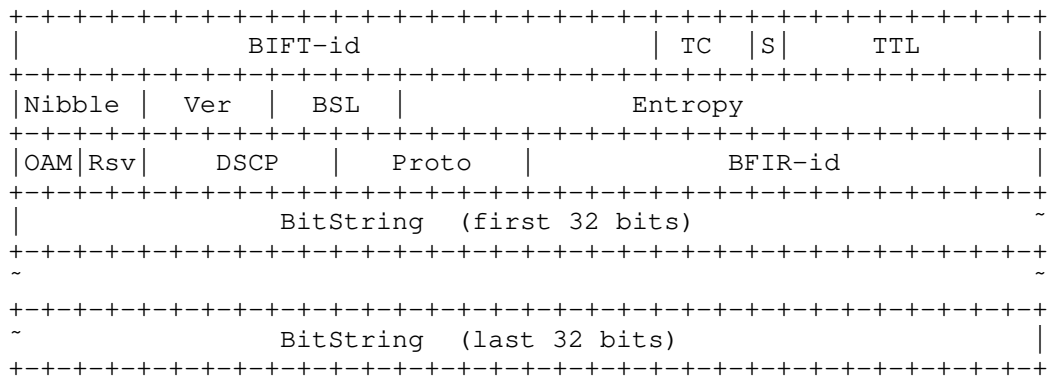
The P-BFIR will become the BFER for multicast traffic point of view. This P-BFIR will track all the PIM interfaces that are interested in the (S,G) and create multicast states for all PIM routers attach to it. This BFER route will have incoming interface (RPF) as BIER "tunnel" interface and outgoing interface as the interface on which PIM Join was received. If there is another PIM Join for the same multicast (S,G) entry on some other interface, that interface gets added in the outgoing interface list.

The P-BFIR after discovering the P-BFER and its BFR-ID (flooded via IGP BIER extension) will construct the BIER header via the BIFT. The PIM packet is encapsulated in the BIER header and transported through BIER domain to P-BFER.

#### 3.1.1. BIER packet construction at PIM BFIR

The BIER header will be encoded with the BFR-id of the P-BFER(with appropriate bit set in the bitstring) and PIM Join is then encapsulated in the packet.

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



BIERHeader.Proto = PIM

BIERHeader.BitString= Bit corresponding to the BFR-ID of the P-BFER

BIERHeader.BFIR-id = BFR-Id of the BER originating the encapsulated PIM packet, i.e. the P-BFIR.

Rest of the values in the BIER header are determined based on the network (MPLS/non-MPLS), capabilities (BSL), and network configuration.

### 3.2. Tunneling PIM through the BIER domain procedure

Throughout the BIER domain the BIER forwarding procedure is in par with draft-ietf-bier-architecture. No BIER router will examine the tunnel PIM packet. As such there is no multicast state build in the BIER domain.

The packet will be forwarded through the BIER domain until it reaches the BER with matching BFR-ID as in the BIERHeader.Bitstring. This BER (P-BFER) will examine the packet and know that it is a PIM packet from BIERHeader.Proto field and farther processing is needed.

### 3.3. PIM-BFER procedure

After receiving the BIER packet and processing the PIM payload encapsulated in BIER packet the P-BFER will remove the BIER header from PIM packet and lookup the route to the source, if the source is in PIM domain, it forwards the PIM packet toward the source.

With same token the P-BFER creates a multicast state with incoming interface as same interface that PIM join packet was forwarded and outgoing interfaces of BIER-Header.BFIR-id.

The P-BFER will also build a BIER reverse path forwarding table (BRT) table, using the BIERHeader.BFIR-id and the Group specified in the arriving PIM packet (S,G). BRT will be used by BFIR for datapath forwarding.

The router keeps track of all BFIR-ids interested in the Group specified in the (S,G) and updates the multicast state and populates the BRT accordingly.

It should be noted this router can also receive and forward PIM packet from other routers in the PIM domain and update the muticast state accordingly.

At this point the end-to-end multicast traffic flow setup is complete.

#### 4. Datapath Forwarding

##### 4.1. BIER reverse path forwarding table

The BIER RPF table (BRT) is needed on the BFIR so the multicast traffic can find the P-BFIR ID and append the correct BIT index to the BIER header for the multicast traffic before forwarding to BIER domain.

This table is built on the P-BFER buy using info from PIM packet and its corresponding BIER header. The PIM packet can provide the specific Group (G) address, meanwhile its corresponding BIER header can provide the originating P-BFIR ID. The P-BFIR is the last BIER router in the BIER domain or the BFER from datapath point of view.

These two pieces of information will be used to build BRT.

It should be noted that a single group can be associated with multiple P-BFIR IDs, as an example multiple MVPN leaf routers behind BIER domain are interested in the same group. These LEAF PEs are reachable via different P-BFIRs.

When the correct P-BFIR(s) (BFER(s)) are found in this table their P-BFIR ID can be used to do a lookup in BIFT and appropriate BIT indexes appended to the BIER header before forwarding the packet to BIER domain.

As an example in the network below:



```

      (BFIR)                                (BFER)
S-- (P-BFER) ----BIER-DOMAIN---- (P-BFIR) --A (join (S,G1), (S,G2))
                                   |-----C (join (S,G1), (S,G3))
                                   |-----E (join (S,G3))

```

The BRT is as follow:

Group	P-BFIR
(G1)	C, A
(G2)	A
(G3)	E, C

And the BIFT is as follow:

BFR-id (SI:Bitstring)	BFR-NBR
1 (0:0001)	C
3 (0:0100)	E
4 (0:1000)	A

As such a multicast dataplain packet arriving with destination G1 will have the BITs (0:1001) and a packet arriving with destination of G3 will have the BITs of (0:0101)

#### 4.2. Datapath traffic flow

When the multicast data traffic arrives on the BFIR (P-BFER) the router will find the destination IP of the traffic (i.e. group address) in the BRT. The BFIR then finds all the P-BFIR (BFER) that are interested in this group from the BRT table. The router then constructs the BIERHeader.BitString with all the BFIR interested in the group and will forward the packet to the BIER domain. The BFER(s) will accept the packets and remove the BIER header and forward the multicast packet as per pre-build multicast state for (G) and its outgoing interfaces.

#### 5. PIM-ASM behavior

In case of PIM ASM the procedure for LEAFs joining RP or the source

is same as above. The unicast (source registration) traffic from source to RP will be flooded throughout the BIER domain as regular unicast traffic without BIER involvement.

## 6. Draft Rosen multicast vpn behavior

Over the years draft rosen mvpn has evolved with many different type of signaling.

As an example, with AD and C-Multicast signaling of PIM or C-Multicast of PIM and AD of BGP.

The above mechanism works with draft rosen mvpn as long the C-multicast signaling is done via PIM. The provider PIM for MVPN can be forwarded from Root and LEAF PE with above explained mechanism.

The multicast traffic has to be forwarded via GRE tunnel. That said the AD signaling can be done via MP-BGP.

Future drafts will address the NG-MVPN and MPLS tunneling.

## 7. Security Considerations

TBD

### 7.1. Normative References

[BIER\_ARCH] Wijnands, IJ., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", internet-draft draft-ietf-bier-architecture-05, October 2016.

[DRAFT ROSEN] E. Rosen, Y. Cai, I. Wijnands, draft-rosen-vpn.mcast-15, June 2010 (RFC 6037)

### 7.2. Informative References

[BGP\_BIER\_EXTENSIONS] Xu, X., Chen, M., Patel, K., Wijnands, I., and A. Przygienda, "BGP Extensions for BIER", internet-draft draft-ietf-bier-idr-extensions-02.txt, June 2017.

[BIER-OAM] Kumar, N., Pignataro, C., Akiya, N., Zheng, L., Chen, M., and G. Mirsky, "BIER Ping and Trace", internet-draft draft-ietf-bier-ping-01.txt, January 2017.

[BIER\_MVPN] Rosen, E., Ed., Sivakumar, M., Wijnands, IJ., Aldrin, S., Dolganow, A., and T. Przygienda, "Multicast VPN Using Bier", internet-draft draft-ietf-bier-mvpn-05, January 2017.

[ISIS\_BIER\_EXTENSIONS] Ginsberg, L., Przygienda, T., Aldrin, S., and Z. Zhang, "BIER Support via ISIS", internet-draft draft-ietf-bier-isis-extensions-04.txt, March 2017.

[OSPF\_BIER\_EXTENSIONS] Psenak, P., Kumar, N., Wijnands, IJ., Dolganow, A., Przygienda, T., Zhang, Z., and S. Aldrin, "OSPF Extensions for Bit Index Explicit Replication", internet-draft draft-ietf-ospf-bier-extensions-05.txt, March 2017.

## 8. Acknowledgments <Add any acknowledgements>

### Authors' Addresses

Hooman Bidgoli (editor)  
Nokia  
600 March Rd.  
Ottawa, Ontario K2K 2E6  
Canada

Email: [hooman.bidgoli@nokia.com](mailto:hooman.bidgoli@nokia.com)

Fengman Xu  
Verizon  
400 International PKWY  
Richardson, Tx 75081  
US

Email: [fengman.xu@verizon.com](mailto:fengman.xu@verizon.com)

Andrew Dolganow  
Nokia  
750D Chai Chee Rd  
06-06, Viva Business Park  
Singapore 469004

Email: [Andrew.dolganow@nokia.com](mailto:Andrew.dolganow@nokia.com)

BIER WG  
Internet-Draft  
Intended status: Standards Track  
Expires: December 29, 2017

Z. Zhang  
ZTE Corporation  
A. Przygienda  
Juniper Networks  
June 27, 2017

BIER in BABEL  
draft-zhang-bier-babel-extensions-01

Abstract

BIER introduces a novel multicast architecture. It does not require a signaling protocol to explicitly build multicast distribution trees, nor does it require intermediate nodes to maintain any per-flow state.

Babel defines a distance-vector routing protocol that operates in a robust and efficient fashion both in wired as well as in wireless mesh networks. This document defines a way to carry necessary BIER signaling information in Babel.

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 <http://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 December 29, 2017.

Copyright Notice

Copyright (c) 2017 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 (<http://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. Terminology . . . . .	2
3. Advertisement of BIER information . . . . .	2
3.1. BIER BFR-prefix and BIER sub-TLV . . . . .	3
3.1.1. BIER sub-TLV . . . . .	3
3.2. BIER MPLS Encapsulation sub-sub-TLV . . . . .	3
3.3. Optional BIER sub-domain BSL conversion sub-sub-TLV . . . . .	4
4. Tree types and tunneling . . . . .	4
5. Security Considerations . . . . .	5
6. IANA Considerations . . . . .	5
7. Acknowledgements . . . . .	5
8. Normative References . . . . .	5
Authors' Addresses . . . . .	6

## 1. Introduction

[I-D.ietf-bier-architecture] introduces a novel multicast architecture. It does not require a signaling protocol to explicitly build multicast distribution trees, nor does it require intermediate nodes to maintain any per-flow state. All procedures necessary to support BIER are abbreviated by the "BIER architecture" moniker in this document.

[RFC6126] and [I-D.ietf-babel-rfc6126bis] define a distance-vector routing protocol under the name of "Babel". Babel operates in a robust and efficient fashion both in ordinary wired as well as in wireless mesh networks.

## 2. Terminology

The terminology of this documents follows  
[I-D.ietf-bier-architecture], [RFC6126], [RFC7557] and  
[I-D.ietf-babel-rfc6126bis].

## 3. Advertisement of BIER information

In case a router is configured with BIER information, and Babel is the routing protocol used, such a router MAY use Babel protocol to announce the BIER information using the BIER sub-TLV specified below.

### 3.1. BIER BFR-prefix and BIER sub-TLV

BFR-prefix and according information is carried in a Babel Update TLV per [I-D.ietf-babel-rfc6126bis]. A new sub-TLV is defined to convey further BIER information such as BFR-id, sub-domain-id and BSL. Two sub-sub-TLVs are carried as payload of BIER sub-TLV.

The mandatory bit of BIER sub-TLV should be set to 0. If a router cannot recognize a sub-TLV, the router MUST ignore this unknown sub-TLV.

#### 3.1.1. BIER sub-TLV

The BIER sub-TLV format aligns exactly with the definition and restrictions in [I-D.ietf-bier-isis-extensions] and [I-D.ietf-bier-ospf-bier-extensions]. It is a sub-TLV of Babel update TLV. The prefix MUST NOT be summarized and the according sub-TLV MUST be treated as optional and transitive.

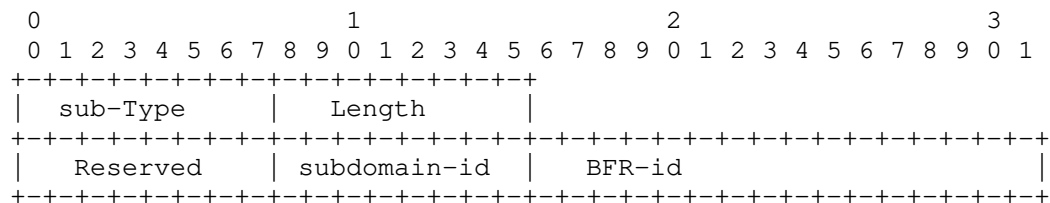


Figure 1: BIER sub-TLV

- o Type: as indicated in IANA section.
- o Length: 1 octet. Include the length of BIER sub-TLV and potential length of the two sub-sub-TLVs.
- o Reserved: MUST be 0 on transmission, ignored on reception. May be used in future versions. 8 bits.
- o subdomain-id: Unique value identifying the BIER sub-domain. 1 octet.
- o BFR-id: A 2 octet field encoding the BFR-id, as documented in [I-D.ietf-bier-architecture]. If no BFR-id has been assigned this field is set to the invalid BFR-id.

### 3.2. BIER MPLS Encapsulation sub-sub-TLV

The BIER MPLS Encapsulation sub-sub-TLV can be carried by BIER sub-TLV. The format and restrictions are aligned with [I-D.ietf-bier-isis-extensions] and

[I-D.ietf-bier-ospf-bier-extensions]. This sub-sub-TLV carries the information for the BIER MPLS encapsulation including the label range for a specific BSL for a certain <MT,SD> pair.

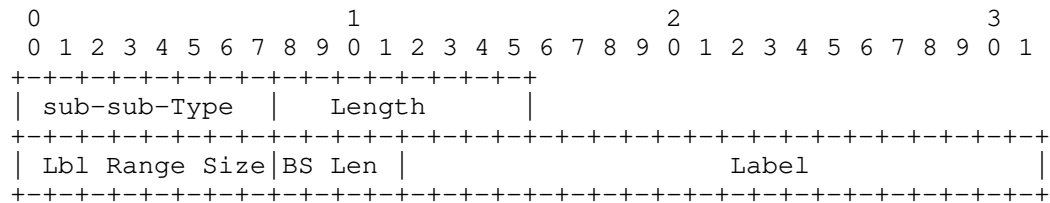


Figure 3: MPLS Encapsulation sub-sub-TLV

- o Type: value of 1 indicating MPLS encapsulation.
- o Length: 1 octet
- o Local BitString Length (BS Len): Encoded bitstring length as per [I-D.ietf-bier-mpls-encapsulation]. 4 bits.
- o Label Range Size: Number of labels in the range for this BIER sub-domain and bitstring length combination, 1 octet.
- o Label: First label of the range, 20 bits. The labels are as defined in [I-D.ietf-bier-mpls-encapsulation].

### 3.3. Optional BIER sub-domain BSL conversion sub-sub-TLV

This sub-sub-TLV is used to carry the BSL information. Its definition and restrictions are aligned with [I-D.ietf-bier-isis-extensions].

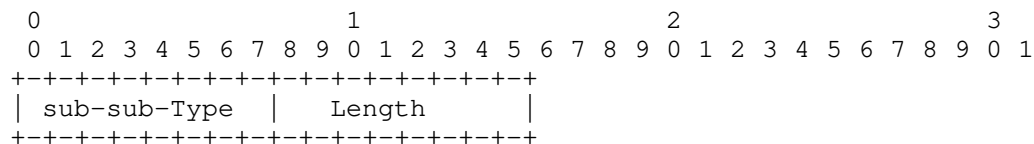


Figure 4: BSL conversion sub-sub-TLV

## 4. Tree types and tunneling

Since Babel is performing a diffusion computation, support for different tree types is not as natural as with link-state protocols. Hence this specification is assuming that normal Babel reachability computation is performed without further modifications.

BIER architecture does not rely on all routers in a domain performing BFR procedures. How to support tunnels that will allow to tunnel BIER across such routers in Babel is for further study.

## 5. Security Considerations

TBD

## 6. IANA Considerations

A new type of Babel update sub-TLV needs to be defined for BIER information advertisement.

## 7. Acknowledgements

The draft is aligned with the [I-D.ietf-bier-isis-extensions] and [I-D.ietf-bier-ospf-bier-extensions] as far as feasible.

## 8. Normative References

[I-D.ietf-babel-rfc6126bis]

Chroboczek, J., "The Babel Routing Protocol", draft-ietf-babel-rfc6126bis-02 (work in progress), May 2017.

[I-D.ietf-bier-architecture]

Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-ietf-bier-architecture-07 (work in progress), June 2017.

[I-D.ietf-bier-isis-extensions]

Ginsberg, L., Przygienda, T., Aldrin, S., and Z. Zhang, "BIER support via ISIS", draft-ietf-bier-isis-extensions-04 (work in progress), March 2017.

[I-D.ietf-bier-mpls-encapsulation]

Wijnands, I., Rosen, E., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication in MPLS and non-MPLS Networks", draft-ietf-bier-mpls-encapsulation-07 (work in progress), June 2017.

[I-D.ietf-bier-ospf-bier-extensions]

Psenak, P., Kumar, N., Wijnands, I., Dolganow, A., Przygienda, T., Zhang, Z., and S. Aldrin, "OSPF Extensions for BIER", draft-ietf-bier-ospf-bier-extensions-06 (work in progress), June 2017.



[RFC6126] Chroboczek, J., "The Babel Routing Protocol", RFC 6126, DOI 10.17487/RFC6126, April 2011, <<http://www.rfc-editor.org/info/rfc6126>>.

[RFC7557] Chroboczek, J., "Extension Mechanism for the Babel Routing Protocol", RFC 7557, DOI 10.17487/RFC7557, May 2015, <<http://www.rfc-editor.org/info/rfc7557>>.

#### Authors' Addresses

Zheng(Sandy) Zhang  
ZTE Corporation  
No. 50 Software Ave, Yuhuatai Distinct  
Nanjing  
China

Email: [zhang.zheng@zte.com.cn](mailto:zhang.zheng@zte.com.cn)

Tony Przygienda  
Juniper Networks

Email: [prz@juniper.net](mailto:prz@juniper.net)

BIER WG  
Internet-Draft  
Intended status: Informational  
Expires: November 25, 2017

Z. Zhang  
B. Wu  
ZTE Corporation  
May 24, 2017

Analysis of BIER in NVO3 network  
draft-zhangwu-bier-nvo3-analysis-00

Abstract

This document analyses BIER deployment in NVO3 network. The intent is to evaluate whether BIER could achieve some simplicity and efficiency.

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 <http://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 November 25, 2017.

Copyright Notice

Copyright (c) 2017 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 (<http://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. BIER in NVO3 network . . . . .	2
2.1. Leaf devices as the edge BIER nodes . . . . .	4
2.2. NVE as edge BIER nodes . . . . .	7
3. Other Consideration . . . . .	8
4. Normative References . . . . .	9
Authors' Addresses . . . . .	10

## 1. Introduction

As mentioned in [I-D.ietf-nvo3-mcast-framework], there is multicast requirement in NVO3 network, such as BUM packets for infrastructure multicast and other application-specific multicast. Besides PIM, BIER is mentioned to be used as one of IP multicast underlay technology in section 3, which introduces several multicast mechanisms for NVO3 network.

Bit Index Explicit Replication (BIER) [I-D.ietf-bier-architecture] is a new architecture for the forwarding of multicast data packets. It provides optimal forwarding of multicast packets through a "multicast domain". It does not require a protocol for explicitly building multicast distribution trees, nor does it require intermediate devices to maintain any per-flow state. When a multicast data packet enters the BIER domain, the BIER ingress router determines the set of BIER egress routers to which the packet needs to be sent. The BIER ingress router then encapsulates the packet in a BIER header. The BIER header contains a bitstring in which each bit represents exactly one BIER egress router in the domain; to forward the packet to a given set of egress routers, the bits corresponding to those routers are set in the BIER header. In this way, elimination of the per-flow state and the explicit tree-building protocols results in a considerable simplification.

This document will give some basic analysis on how to deploy BIER in the NVO3 network to mitigate multicast states.

## 2. BIER in NVO3 network

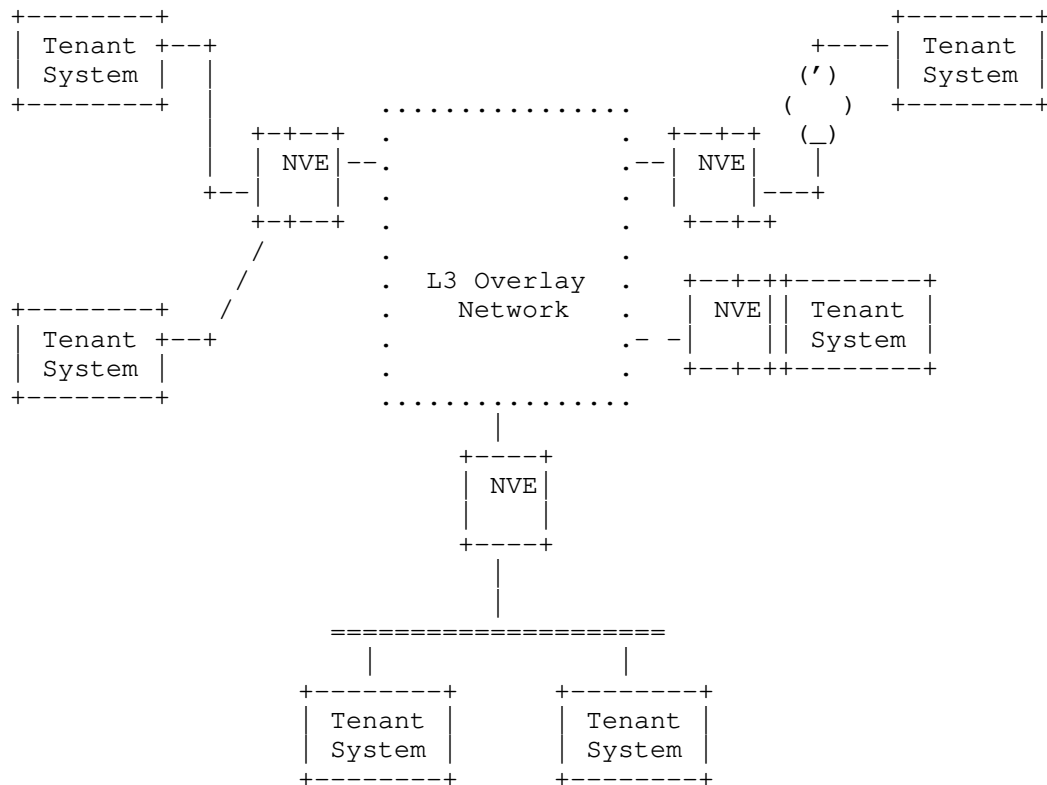


Figure 1: NVO3 Generic Reference Model

As described in [RFC8014], figure 1 is a generic reference model of NVO3. BIER can be used in this model to provide underlay multicast function.

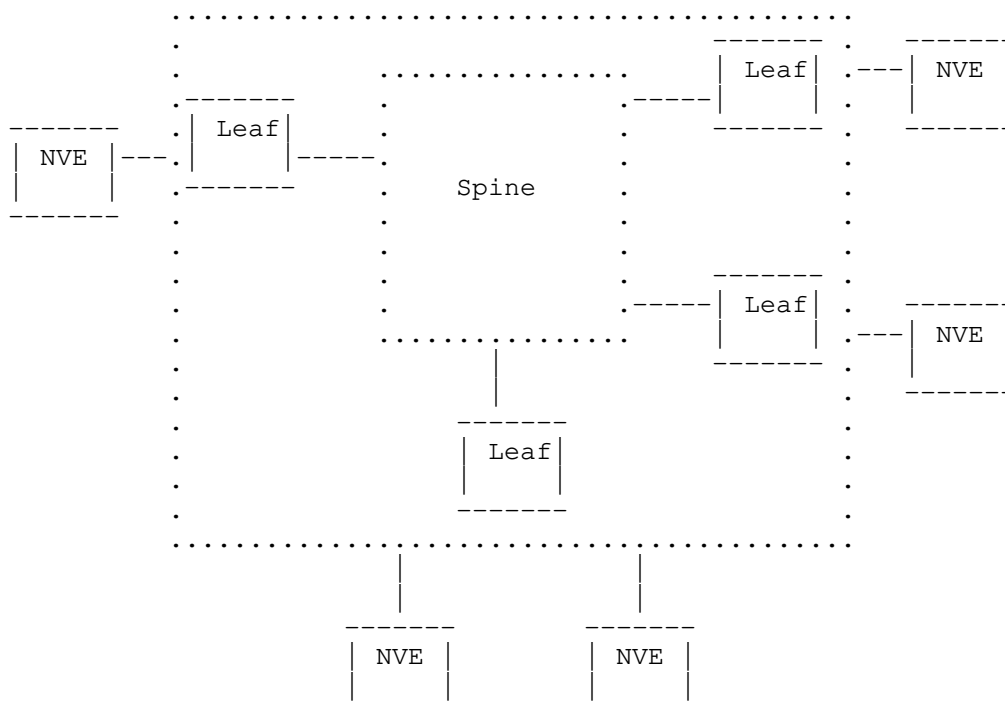


Figure 2: Example topology of NV03 underlay network

The architecture of L3 underlay Network can be various. As depicted in figure 2, the Clos topology mentioned in [RFC7938] is taken as an example to offer underlay L3 connectivity. And it is assumed that the leaf and spine devices support BIER functionality including exchanging BFR-id and other information and building BIER forwarding table.

In this document, two possible ways to handle multicast packets are identified. One of the choices is placing BIER boundaries at leaf switches which are a general choice of implementing BIER. The other way is putting the BIER boundary at NVE to achieve better efficiency.

### 2.1. Leaf devices as the edge BIER nodes

Leaf devices are taken as BIER edge node. Each leaf device is allocated one unique BFR-id. And leaf devices are the ingress BIER nodes and egress BIER nodes. Both the leaf and spine devices exchange the BIER information by IGP/BGP extensions. The according extensions are defined in [I-D.ietf-bier-ospf-bier-extensions], [I-D.ietf-bier-isis-extensions], and [I-D.ietf-bier-idr-extensions].

IGMP/MLD protocol is used between NVE and leaf device like the description in [I-D.ietf-nvo3-mcast-framework]. The BUM flows are encapsulated corresponding multicast group address. BMLD [I-D.pfister-bier-mld] protocol is used among all the leaf devices to exchange multicast group information. After BMLD protocol process is completed, each leaf devices knows the other leaf devices associated with specific multicast group address. When one packet with multicast group address reaches leaf device, leaf device encapsulates the packet with BIER header which indicates all the destination leaf devices that belong to the same multicast group.

After the BIER packet reaches the destination leaf devices through the spine network forwarding, the destination leaf device removes the BIER header of packet and forwards to corresponding NVEs.

So the multicast state is eliminated because of removing of multicast tree in L3 underlay network. NVE only needs to support IGMP/MLD protocol. But one or several multicast group addresses for a tenant is still needed.



Leaf1 send two copies of packet to Leaf2 and Leaf3. SI in BIER header defined in [I-D.ietf-bier-architecture] is be used to distinguish these two copies to deliver to the final destination. The other solution is increase the forwarding BitStringLength to 128. From the above analysis, BIER is quite applicable in the scenario with limited size of leaf switches. But in a large scale of NVO3 underlay network, there is some limitation due to the BitStringLength.

## 2.2. NVE as edge BIER nodes

In last section, IGMP/MLD is still needed to run between NVE and leaf devices. And the multicast groups for Tenants and specific multicast applications are needed.

If NVE is the edge BIER node, IGMP/MLD protocol does not need to run between NVE and leaf device. NVE encapsulates the BIER header with the BFR-ids of destination NVEs straightly and send it to leaf devices. Leaf and spine devices forward the BIER packet to destination NVEs. Destination NVEs remove the BIER header and forwarding according to the inner encapsulation.

In this situation, leaf device does not need to be allocated BFR-id. Every NVE should be allocated with one unique BFR-id. The BFR-id information should be exchanged within the L3 underlay network. If NVE supports the IGP/BGP BIER extension, NVE takes part in the BIER information exchange. The forwarding plane will be established easily. But in some situations NVE does not support IGP or BGP; NVE can not take part in the BIER information exchange. So the BFR-id of NVE should be advertised by some other method.

Besides multicast state elimination in L3 underlay network, IGMP/MLD does not need to run between NVE and leaf devices. And BMLD does not need to run among all the leaf devices. This function makes the BIER deployment more simply.



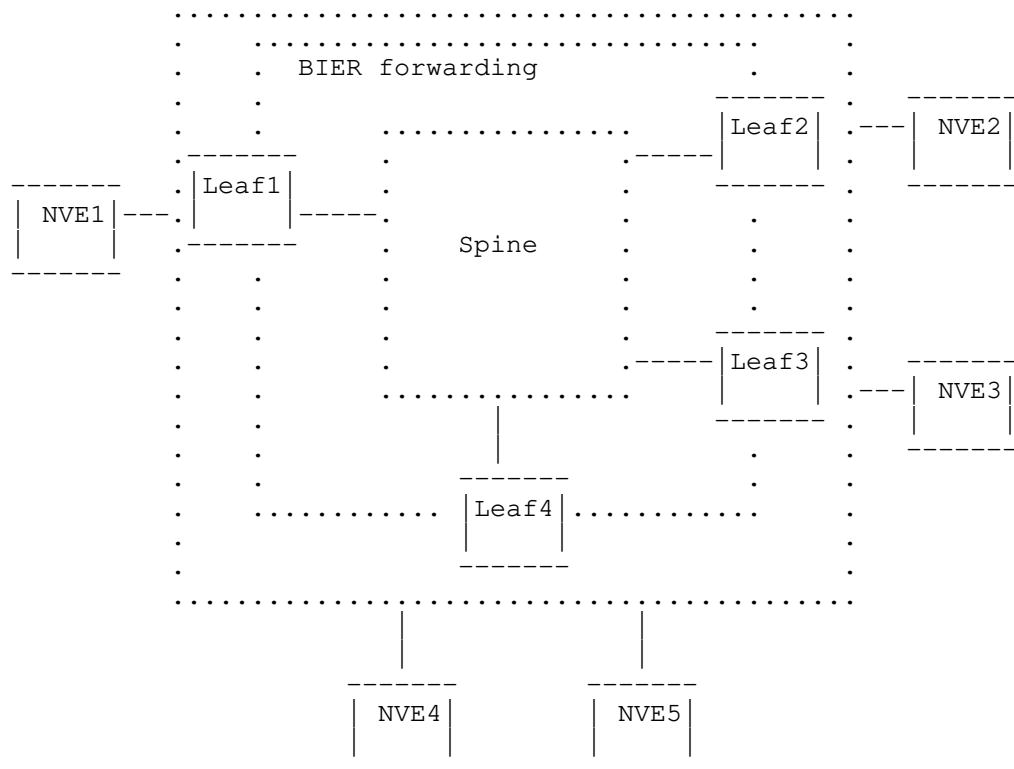


Figure 4: BIER in VNO3

The same example for figure 3, NVE1 should send BUM packet to NVE2, NVE3, NVE4 and NVE5. NVE1 encapsulates BIER header with NVE2, NVE3, NVE4 and NVE5 as destination and send it to Leaf1. Leaf1 and spine devices forward the packet to NVE2, NVE3, NVE4 and NVE5 according to the BIER header. The NVEs remove the BIER header and forward it.

The BitStringLength limitation also remains in this solution. And the situation may be more seriously because of the larger number of NVEs than leaf devices. If the BFR-id of NVE is not allocated reasonably, in the worst situation, the forwarding efficiency is the same with the source replication described in [I-D.ietf-nvo3-mcast-framework] section 3.2.

### 3. Other Consideration

As illustrated in [RFC7938], there could be hundred thousand servers connected by underlay network in some NVO3 network. So there are more than thousands of NVEs and leaf devices in the network. Using BIER as multicast underlay protocol make significant advantage

because of the elimination of multicast state stored in the underlay network. But the BitStringLength limitation is one problem.

In order to achieve the optimization of BIER, the BFR-ids allocation should be more reasonable. The BFR-id of NVE/leaf device that is belong to one same VN could be allocated adjacent as much as possible. So the encapsulation of BIER header can be more efficient.

Along with the number of BFR-id increasing for NVE/leaf devices, there are thousands BIER forwarding items in the L3 underlay network. The forwarding efficiency in the L3 underlay network should also be considered.

#### 4. Normative References

[I-D.ietf-bier-architecture]

Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-ietf-bier-architecture-06 (work in progress), April 2017.

[I-D.ietf-bier-idr-extensions]

Xu, X., Chen, M., Patel, K., Wijnands, I., and T. Przygienda, "BGP Extensions for BIER", draft-ietf-bier-idr-extensions-02 (work in progress), January 2017.

[I-D.ietf-bier-isis-extensions]

Ginsberg, L., Przygienda, T., Aldrin, S., and Z. Zhang, "BIER support via ISIS", draft-ietf-bier-isis-extensions-04 (work in progress), March 2017.

[I-D.ietf-bier-ospf-bier-extensions]

Psenak, P., Kumar, N., Wijnands, I., Dolganow, A., Przygienda, T., Zhang, Z., and S. Aldrin, "OSPF Extensions for BIER", draft-ietf-bier-ospf-bier-extensions-05 (work in progress), March 2017.

[I-D.ietf-nvo3-mcast-framework]

Ghanwani, A., Dunbar, L., McBride, M., Bannai, V., and R. Krishnan, "A Framework for Multicast in Network Virtualization Overlays", draft-ietf-nvo3-mcast-framework-08 (work in progress), May 2017.

[I-D.pfister-bier-mld]

Pfister, P., Wijnands, I., Venaas, S., Wang, C., Zhang, Z., and M. Stenberg, "BIER Ingress Multicast Flow Overlay using Multicast Listener Discovery Protocols", draft-pfister-bier-mld-03 (work in progress), March 2017.

- [RFC7938] Lapukhov, P., Premji, A., and J. Mitchell, Ed., "Use of BGP for Routing in Large-Scale Data Centers", RFC 7938, DOI 10.17487/RFC7938, August 2016, <<http://www.rfc-editor.org/info/rfc7938>>.
- [RFC8014] Black, D., Hudson, J., Kreeger, L., Lasserre, M., and T. Narten, "An Architecture for Data-Center Network Virtualization over Layer 3 (NVO3)", RFC 8014, DOI 10.17487/RFC8014, December 2016, <<http://www.rfc-editor.org/info/rfc8014>>.

## Authors' Addresses

Zheng(Sandy) Zhang  
ZTE Corporation  
No. 50 Software Ave, Yuhuatai District  
Nanjing  
China

Email: [zhang.zheng@zte.com.cn](mailto:zhang.zheng@zte.com.cn)

Bo Wu  
ZTE Corporation  
No. 50 Software Ave, Yuhuatai District  
Nanjing  
China

Email: [wu.bo@zte.com.cn](mailto:wu.bo@zte.com.cn)

BIER  
Internet-Draft  
Intended status: Standards Track  
Expires: December 10, 2017

Z. Zhang  
A. Przygienda  
Juniper Networks  
A. Sajassi  
Cisco Systems  
J. Rabadan  
Nokia  
June 8, 2017

EVPN BUM Using BIER  
draft-zzhang-bier-evpn-00

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 <http://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 December 10, 2017.

Copyright Notice

Copyright (c) 2017 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 (<http://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 . . . . .	6
3. Multihoming Split Horizon . . . . .	7
4. Data Plane . . . . .	8
4.1. Encapsulation and Transmission . . . . .	8
4.2. Disposition . . . . .	9
4.2.1. At a BFER that is an Egress PE . . . . .	9
4.2.2. At a BFER that is a P-tunnel Segmentation Boundary . . . . .	10
5. IANA Considerations . . . . .	10
6. Security Considerations . . . . .	10
7. Acknowledgements . . . . .	10
8. References . . . . .	10
8.1. Normative References . . . . .	10
8.2. Informative References . . . . .	11
Authors' Addresses . . . . .	11

## 1. Introduction

[RFC7432] and [I-D.ietf-bess-evpn-overlay] 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) ([I-D.ietf-bier-architecture]) 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.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 that [I-D.ietf-bier-mvpn] requests IANA to assign 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 the BFR-Prefix (see [I-D.ietf-bier-architecture]) 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.
- 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 [I-D.ietf-bess-evpn-overlay], 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 [I-D.ietf-bier-architecture]). 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 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 LIR bit may be set in the S-PMSI A-D routes, and the PEs that need to receive corresponding traffic will respond with a Leaf A-D route. The ingress PE identifies the set of BFERs to deliver traffic to according to the set of corresponding Leaf A-D routes received.

The S-PMSI A-D route carries the same PTA as in the IMET route, except that similar to MVPN, the LIR-pF flag may be set for an ingress PE to request individual (C-S,C-G) or (C-\*,C-G) Leaf A-D routes.

#### 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.zhang-bess-mvpn-evpn-cmcast-enhancements] for segmentation with SMET routes. This is only applicable to EVPN-MPLS.

Similar to MVPN, the LIR-pF flag cannot be used with segmentation, and the S-PMSI A-D routes' PTA MUST carry an upstream assigned label to allow tunnel segmentation points to do label switching. The S-PMSI A-D routes could be proactively (re-)advertised by the ingress PEs or segmentation points, or could be triggered by the unsolicited Leaf A-D routes received from downstream.

## 2.2. MPLS Label in PTA

Similar to the MVPN case in [I-D.ietf-bier-mvpn], the label allocation for the upstream assigned label in the PTA MUST follow the following rules (text borrowed verbatim from [I-D.ietf-bier-mvpn]).

Suppose an ingress PE originates two x-PMSI A-D routes, where we use the term "x-PMSI" to mean "I-PMSI or S-PMSI". Suppose both routes carry a PTA, and the PTA of each route specifies "BIER".

- o If the two routes do not carry the same set of Route Targets (RTs), then their respective PTAs MUST contain different MPLS label values.
- o If segmented P-tunnels are being used, then the respective PTAs of the two routes MUST contain different MPLS label values, as long as the NLRIs are not identical. In this case, the MPLS label can be used by the BFER to identify the particular C-flow to which a data packet belongs, and this greatly simplifies the process of forwarding a received packet to its next P-tunnel segment. This is explained further below.

When segmented P-tunnels are being used, an ABR or ASBR may receive, from a BIER domain, an x-PMSI A-D route whose PTA specifies "BIER". This means that BIER is being used for one segment of a segmented P-tunnel. The ABR/ASBR may in turn need to originate an x-PMSI A-D route whose PTA identifies the next segment of the P-tunnel. The next segment may also be "BIER". Suppose an ABR/ASBR receives x-PMSI A-D routes R1 and R2, and as a result originates x-PMSI A-D routes R3 and R4 respectively, where the PTAs of each of the four routes specify BIER. Then the PTAs of R3 and R4 MUST NOT specify the same MPLS label.

The ABR/ASBR MUST then program its dataplane such that a packet arriving with the upstream-assigned label specified in route R1 is transmitted with the upstream-assigned label specified in route R3, and a packet arriving with the upstream-assigned label specified in route R2 is transmitted with the label specified in route R4. Of course, the data plane must also be programmed to encapsulate the transmitted packets with an appropriate BIER header, whose BitString is determined by the multicast flow overlay.

### 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, [I-D.ietf-bess-evpn-overlay] specifies local-bias procedures, where 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.

Details for split-horizon in case of segmentation will be provided in future revisions.

#### 4. Data Plane

Similar to MVPN, the EVPN application plays the role of the "multicast flow overlay" as described in [I-D.ietf-bier-architecture].

##### 4.1. Encapsulation and Transmission

To transmit a BUM data packet, an ingress PE first pushes the 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.

It then finds the S-PMSI A-D route, or the SMET/IMET route that matches that packet. Any S-PMSI A-D route with a PTA specifying "no tunnel information" is ignored. If one or more SMET routes are matched, the IMET route originated by the ingress PE for the broadcast domain is then located to obtain the PTA.

If the found S-PMSI A-D or the IMET route has a PTA specifying "BIER", and the ingress PE determines that BIER should be used (e.g., per procedures in [I-D.ietf-bess-evpn-igmp-mld-proxy] about interworking with PEs that do not support certain tunnel types), the (upstream-assigned) MPLS label from that PTA is pushed on 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 [I-D.ietf-bier-architecture] and [I-D.ietf-bier-mpls-encapsulation]. See especially Section 4, "Imposing and Processing the BIER Encapsulation", of [I-D.ietf-bier-mpls-encapsulation]. 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. If SMET routes are matched, it determines all the BFERs from all the matching SMET routes in the broadcast domain.

If an S-PMSI route is matched, it determines all the BFERs by finding all the Leaf A-D routes that correspond to the S-PMSI A-D route that is the packet's match for transmission. There are two different cases to consider:

- 1 The S-PMSI A-D route that is the match for transmission carries a PTA that has the LIR flag set but does not have the LIR-pF flag set. In this case, the corresponding Leaf A-D routes are those whose "route key" field is identical to the NLRI of the S-PMSI A-D route.
- 2 The S-PMSI A-D route that is the match for transmission carries a PTA that has the LIR-pF flag. In this case, the corresponding Leaf A-D routes are those 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 [EXPLICIT\_TRACKING].

#### 4.2. Disposition

The same procedures in section 3.2 of [I-D.ietf-bier-mvpn] are followed for EVPN-MPLS (text could be copied here). 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 [I-D.ietf-bess-evpn-overlay] 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 Boundary

This is only applicable to EVPN-MPLS. The same procedures in Section 3.2.2 of [I-D.ietf-bier-mvpn] are followed, subject to multihoming considerations described in Section 3 of this document.

### 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., and K. Patel, "Updates on EVPN BUM Procedures", draft-ietf-bess-evpn-bum-procedure-updates-01 (work in progress), December 2016.
- [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-00 (work in progress), March 2017.
- [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-02 (work in progress), June 2017.

- [I-D.ietf-bier-architecture]  
Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-ietf-bier-architecture-06 (work in progress), April 2017.
- [I-D.ietf-bier-mpls-encapsulation]  
Wijnands, I., Rosen, E., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication in MPLS and non-MPLS Networks", draft-ietf-bier-mpls-encapsulation-07 (work in progress), June 2017.
- [I-D.ietf-bier-mvpn]  
Rosen, E., Sivakumar, M., Aldrin, S., Dolganow, A., and T. Przygienda, "Multicast VPN Using BIER", draft-ietf-bier-mvpn-05 (work in progress), January 2017.
- [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.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [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, <<http://www.rfc-editor.org/info/rfc7432>>.

## 8.2. Informative References

- [I-D.ietf-bess-evpn-overlay]  
Sajassi, A., Drake, J., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution using EVPN", draft-ietf-bess-evpn-overlay-08 (work in progress), March 2017.

### Authors' Addresses

Zhaohui Zhang  
Juniper Networks

EMail: [zzhang@juniper.net](mailto: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