Workgroup: Network Working Group
Internet-Draft:
draft-ietf-bier-idr-extensions-09
Published: 15 February 2023
Intended Status: Standards Track
Expires: 19 August 2023
Authors: X.X. Xu
Midea Group

M.C. Chen
Huawei

K.P. Patel
Arrcus, Inc.

I.W. Wijnands
Individual

A.P. Przygienda
Juniper

Z. Zhang, Ed.
Juniper

BGP Extensions for BIER

Abstract

Bit Index Explicit Replication (BIER) is a new multicast forwarding architecture which doesn't require an explicit tree-building protocol and doesn't require intermediate routers to maintain any multicast state. BIER is applicable in a multi-tenant data center network environment for efficient delivery of Broadcast, Unknown-unicast and Multicast (BUM) traffic while eliminating the need for maintaining a huge amount of multicast state in the underlay. This document describes BGP extensions for advertising the BIER-specific information.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 19 August 2023.

Table of Contents

1. Introduction

Bit Index Explicit Replication (BIER) [RFC8279] is a new multicast forwarding architecture which doesn't require an explicit tree-building protocol and doesn't require intermediate routers to maintain any multicast state. BIER is applicable in a multi-tenant data center network environment for efficient delivery of Broadcast, Unknown-unicast and Multicast (BUM) traffic while eliminating the need for maintaining a huge amount of multicast state in the underlay. This document describes BGP extensions for advertising the BIER-specific information. More specifically, in this document, we define a new optional, non- transitive BGP attribute, referred to as the BIER attribute, to convey the BIER-specific information such as BIER Forwarding Router identifier (BFR-id), BitString Length (BSL)

and so on. In addition, this document specifies procedures to prevent the BIER attribute from "leaking out" of a BIER domain.

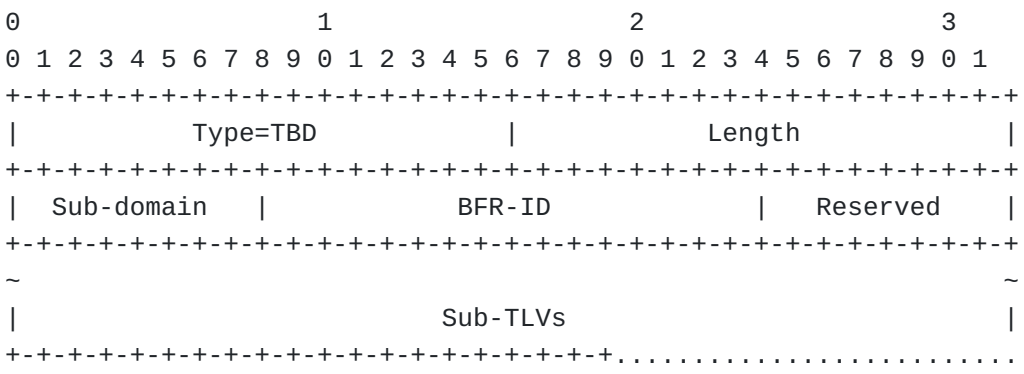
These extensions are applicable in those multi-tenant data centers where BGP instead of IGP is used as an underlay [RFC7938]. These extensions may also be applicable to other BGP based network scenarios, e.g., as described in [I-D.ietf-bier-multicast-as-a-service].

2. Terminology

This memo makes use of the terms defined in [RFC4271] and [RFC8279].

3. BIER Path Attribute

This draft defines a new optional, transitive BGP path attribute, referred to as the BIER attribute. This attribute can be attached to a BGP UPDATE message by the originator so as to indicate the BIER-specific information of a particular BFR which is identified by the /32 or /128 address prefix contained in the NLRI. In other words, if the BIER path attribute is present, the NLRI is treated by BIER as a "BFR-prefix". When creating a BIER attribute, a BFR needs to include one BIER TLV for every Sub-domain that it supports. The attribute type code for the BIER Attribute is TBD. The value field of the BIER Attribute contains one or more BIER TLV as shown in Figure 1.



Type: Two octets encoding the BIER TLV Type: TBD.

Length: Two octets encoding the length in octets of the TLV, including the type and length fields. The length is encoded as an unsigned binary integer. (Note that the minimum length is 8, indicating that no sub-TLV is present.)

Sub-domain: a one-octet field encoding the sub-domain ID corresponding to the BFR-ID.

BFR-ID: a two-octet field encoding the BFR-ID.

Sub-TLVs: contains one or more sub-TLV.

The BIER TLV MAY appear multiple times in the BIER Path Attribute, one for each sub-domain. There MUST be no more than one BIER TLV with the same Sub-domain value; if there is, the entire BIER Path Attribute MUST be ignored.

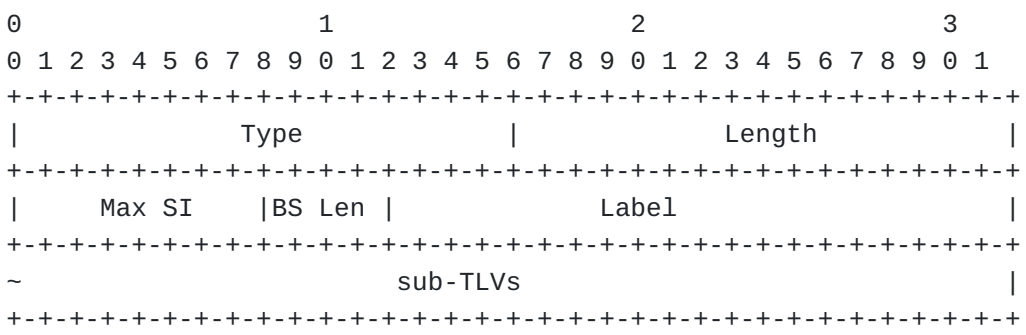
A BIER TLV may have sub-TLVs, which may have their own sub-TLVs. All those are referred to as sub-TLVs and share the same Type space, regardless of the level.

3.1. BIER MPLS Encapsulation sub-TLV

The BIER MPLS Encapsulation sub-TLV matches the OSPFv2 "BIER MPLS Encapsulation sub-TLV" as specified in Section 2.2 of [RFC8444]. It MAY appear multiple times in the BIER TLV.

The following is copied verbatim from that section:

The BIER MPLS Encapsulation Sub-TLV has the following format:



Type: TBD1 (To be assigned by IANA).

Length: 4 or other values (depending on sub-TLVs)

Max SI: A 1-octet field encoding the maximum Set Identifier (SI) (see Section 1 of [RFC8279]) used in the encapsulation for this BIER sub-domain for this BitString length.

BS Len (BitString Length): A 4-bit field encoding the supported BitString length associated with this BFR-prefix. The values allowed in this field are specified in Section 2 of [RFC8296].

Label: A 20-bit value representing the first label in the label rang

The "label range" is the set of labels beginning with the Label and ending with (Label + (Max SI)). A unique label range is allocated for each BitString length and sub-domain-id. These labels are used for BIER forwarding as described in [RFC8279] and [RFC8296].

The size of the label range is determined by the number of SIs (Section 1 of [RFC8279]) that are used in the network. Each SI maps to a single label in the label range: the first label is for SI=0, the second label is for SI=1, etc.

If the label associated with the Maximum Set Identifier exceeds the 20-bit range, the BIER MPLS Encapsulation Sub-TLV containing the error MUST be ignored.

If the same BitString length is repeated in multiple BIER MPLS Encapsulation Sub-TLVs inside the same BIER TLV, the whole BIER TLV containing the conflicts MUST be ignored.

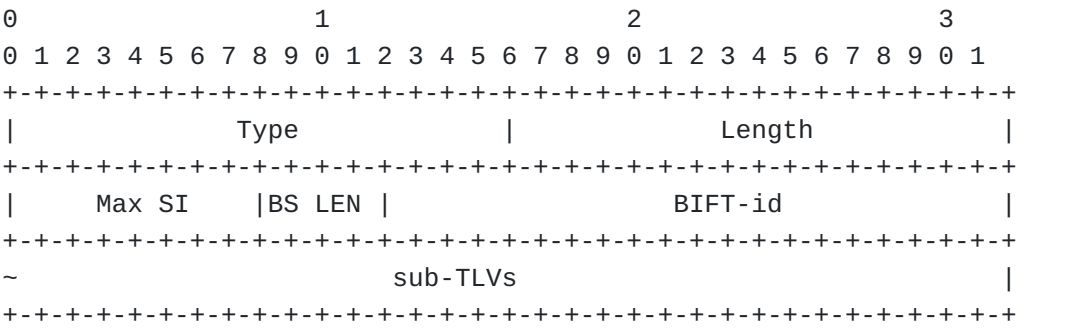
Label ranges within all BIER MPLS Encapsulation Sub-TLVs advertised by the same BFR MUST NOT overlap. If an overlap is detected, all BIER TLVs advertised by such a router MUST be ignored.

3.2. BIER Non-MPLS Encapsulation sub-TLV

Similar to the concept in [I-D.ietf-bier-lsr-non-mpls-extensions], the BIER non-MPLS Encapsulation sub-TLV is used for non-MPLS encapsulation. It matches the OSPFv2 BIER non-MPLS Encapsulation sub TLV as specified in Section 3.2 of [I-D.ietf-bier-lsr-non-mpls-extensions].

The following are copied verbatim from that section. Note to RFC Editor: the following copied text must match the final text in the RFC for [I-D.ietf-bier-lsr-non-mpls-extensions].

The non-MPLS Encapsulation Sub-TLV MAY appear multiple times within a single BIER TLV. If the same BitString length is repeated in multiple BIER non-MPLS encapsulation Sub-TLVs inside the same BIER TLV, the BIER TLV MUST be ignored.



Type: TBD2 (To be assigned by IANA).

Length: 4 or other values (depending on sub-TLVs)

Max SI: A 1 octet field encoding the Maximum Set Identifier (Section 1 of [RFC8279]) used in the encapsulation for this BIER subdomain for this BitString length. The first BIFT-id is for SI= the second BIFT-id is for SI=1, etc. If the BIFT-id associated wi the Maximum Set Identifier exceeds the 20-bit range, the sub-TLV MUST be ignored.

BIFT-id: A 20-bit field representing the first BIFT-id in the BIFT-i range.

BitString Length (BS Len): A 4 bit field encoding the bitstring length (as per [RFC8296]) supported for the encapsulatio

The "BIFT-id range" is the set of 20-bit values beginning with the BIFT-id and ending with (BIFT-id + (Max SI)). These BIFT-id's are used for BIER forwarding as described in [RFC8279] and [RFC8296].

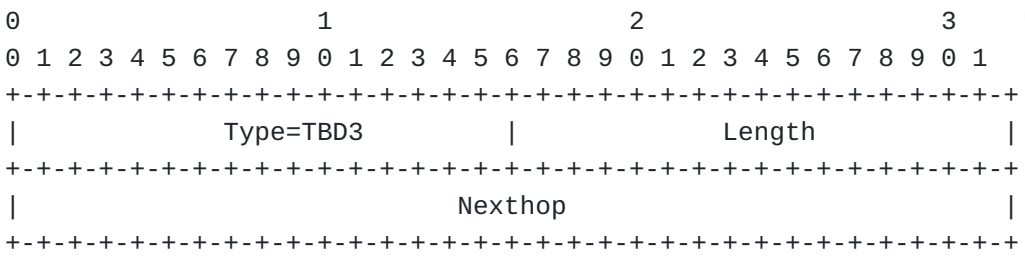
The size of the BIFT-id range is determined by the number of SI's (Section 1 of [RFC8279]) that are used in the network. Each SI maps to a single BIFT-id in the BIFT-id range: the first BIFT-id is for SI=0, the second BIFT-id is for SI=1, etc.

If the BIFT-id associated with the Maximum Set Identifier exceeds the 20-bit range, the BIER non-MPLS Encapsulation sub-TLV containing the error MUST be ignored.

BIFT-id ranges within all the BIER non-MPLS Encapsulation sub-TLVs advertised by the same BFR MUST NOT overlap. If the overlap is detected, the advertising router MUST be treated as if it did not advertise any BIER non-MPLS encapsulation sub-TLVs. However the

BIFT-id ranges may overlap across different encapsulation types and is allowed. As an example, the BIFT-id value in the non-MPLS encapsulation sub-TLV may overlap with the Label value in the Label range in BIER MPLS encapsulation sub-TLV.

3.3. BIER Nexthop sub-TLV



Type: TBD3 (To be assigned by IANA).

Length: 4 if the Nexthop is IPv4 address and 16 if the Nexthop is IPv6 address

Nexthop: 4 or 16 octets of IPv4/IPv6 address

The BIER Nexthop sub-TLV MAY be included in the MPLS or non-MPLS Encapsulation sub-TLV as well as in the top level BIER TLV.

4. Originating/Updating BIER Attribute

A BIER Forwarding Egress Router (BFER) MUST attach a BIER attribute to its own BIER prefix NLRI. The BIER attribute MUST include one BIER TLV for each BIER sub-domain that it supports. Each BIER TLV MUST include an MPLS and/or non-MPLS Encapsulation sub-TLV, and SHOULD include a BIER Nexthop sub-TLV with the Nexthop set to the BIER prefix. If the BIER Nexthop sub-TLV is not included, the BIER prefix will be used by receiving BFRs as the BIER nexthop when calculating BIFT.

A BFR/BFER MAY attach a BIER proxy range sub-TLV [I-D.ietf-bier-prefix-redistribute] in the BIER TLV. In this case it MUST attach a BIER attribute to its own BIER prefix NLRIs. Other than this case, a BFR that is not a BFER (i.e., its BFR-ID is 0) SHOULD NOT attach a BIER attribute to its own BIER prefix NLRIs (if a BIER attribute is attached it will not get used anyway).

When a BFR re-advertises a BGP NLRI with a BIER attribute, it SHOULD set/update the BIER Nexthop sub-TLV to use its own BIER prefix, in which case it MUST replace the MPLS or non-MPLS Encapsulation sub-TLV with its own, i.e., as if the BFR is attaching the encapsulation sub-TLV for its own BIER prefix. If it does not update the BIER Nexthop sub-TLVs, it MUST NOT update MPLS or non-MPLS Encapsulation sub-TLV.

It's possible that the BFR supports some but not all BSLs in the received MPLS or non-MPLS Encapsulation sub-TLVs. After updating the BIER Nexthop sub-TLV in the top BIER TLV to itself, for the BSLs

that it does support, the BFR MUST remove the BIER Nexthop sub-TLV (if present) in the corresponding Encapsulation sub-TLVs. For the BSLs that it does not support, it MUST not update those Encapsulation sub-TLVs except that if a BIER Nexthop sub-TLV is not included in the Encapsulation sub-TLV, the received BIER Nexthop sub-TLV in the top BIER TLV MUST be copied into the Encapsulation sub-TLV. All impacted length fields (e.g., the Encapsulation sub-TLV Length, the top level BIER TLV Length) MUST be updated accordingly.

Since the BIER attribute is an optional, transitive BGP path attribute, a non-BFR BGP speaker could still advertise the received route with a BIER attribute.

5. BIFT Calculation

For each sub-domain, a BFR calculates the corresponding BIFTs by going through the BIER prefixes whose BIER attribute includes a BIER TLV for the sub-domain. For a non-zero BFR-id in the BIER TLV, or for each BFR-id in the BIER Proxy Range sub-TLV in the BIER TLV of a BIER prefix, a BIFT entry is created or updated. The entry's BFR Neighbor (BFR-NBR) [RFC8279] is the Nexthop in the BIER Nexthop sub-TLV in the corresponding Encapsulation sub-TLV, or in the top level BIER TLV if the Encapsulation sub-TLV does not have a Nexthop sub-TLV. If there is no Nexthop sub-TLV at all, The entry's BFR Neighbor is the BIER prefix itself. The BIER label or BIFT-id for the entry is derived from the Label Range in the MPLS Encapsulation sub-TLV or from the BIFT-id Range in the non-MPLS Encapsulation sub-TLV.

BIER traffic is sent to the BFR-NBR either natively (BIER header directly follows a layer 2 header) if the BFR-NBR is directly connected, or via a tunnel otherwise. Notice that, if a non-BFR BGP speaker re-advertises a BIER prefix (in this case it can not update the BIER attribute since it is not capable), or if a BFR BGP speaker re-advertises a BIER prefix without updating the BIER Nexthop sub-TLV, the BFR receiving the prefix will tunnel BIER traffic - the BGP speaker re-advertising the BIER prefix will not see the BIER traffic for the BIER prefix.

6. Deployment Considerations

It's assumed by this document that the BIER domain is aligned with an Administrative Domain (AD) which may be composed of multiple ASes (either private or public ASes). Use of the BIER attribute in other scenarios is outside the scope of this document.

A boundary router of the AD that supports the BIER attribute MUST support a per-EBGP-session/group policy, that indicates whether the attribute is allowed. If it is not allowed, the BIER attribute MUST NOT be sent to any EBGP peer of the session/group, and the BIER

attribute received from the peer MUST be treated exactly as if it were an unrecognized non-transitive attribute. That is, "it MUST be quietly ignored and not passed along to other BGP peers".

7. Acknowledgements

Thanks a lot for Eric Rosen and Peter Psenak for their valuable comments on this document.

8. IANA Considerations

IANA is requested to assign a codepoint in the "BGP Path Attributes" registry to the BIER attribute.

IANA is requested to create a registry for "BGP BIER Attribute Types" and a registry for "BGP BIER TLV sub-TLV Types". The type field for both registry consists of two octets, with possible values from 1 to 655355 (the value 0 is reserved). The allocation policy for this field is to be "First Come First Serve".

Three initial values are to be allocated from the "BGP BIER TLV sub-TLV Types" for MPLS Encapsulation, non-MPLS Encapsulation, and BIER Nexthop sub-TLV respectively.

9. Security Considerations

This document introduces no new security considerations beyond those already specified in [RFC4271] and [RFC8279].

10. Contributors

This document has the following contributors:

Zheng Zhang
ZTE
zhang.zheng@zte.com.cn

11. References

11.1. Normative References

[I-D.ietf-bier-lsr-non-mpls-extensions]
Dhanaraj, S., Yan, G., Wijnands, I., Psenak, P., Zhang, Z. J., and J. Xie, "LSR Extensions for BIER non-MPLS Encapsulation", Work in Progress, Internet-Draft, draft-ietf-bier-lsr-non-mpls-extensions-01, 19 September 2022,

<https://www.ietf.org/archive/id/draft-ietf-bier-lsr-non-mpls-extensions-01.txt>.

[I-D.ietf-bier-prefix-redistribute]
Zhang, Z., Wu, B., Zhang, Z. J., Wijnands, I., Liu, Y., and H. Bidgoli, "BIER Prefix Redistribute", Work in Progress, Internet-Draft, draft-ietf-bier-prefix-redistribute-03, 3 January 2023, <https://www.ietf.org/archive/id/draft-ietf-bier-prefix-redistribute-03.txt>.

[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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <https://www.rfc-editor.org/rfc/rfc8279>.

[RFC8444] Psenak, P., Ed., Kumar, N., Wijnands, IJ., Dolganow, A., Przygienda, T., Zhang, J., and S. Aldrin, "OSPFv2 Extensions for Bit Index Explicit Replication (BIER)", RFC 8444, DOI 10.17487/RFC8444, November 2018, <https://www.rfc-editor.org/info/rfc8444>.

11.2. Informative References

[I-D.ietf-bier-multicast-as-a-service]
Zhang, Z. J., Rosen, E. C., Awduche, D. O., Shepherd, G., Zhang, Z., and G. S. Mishra, "Multicast/BIER As A Service", Work in Progress, Internet-Draft, draft-ietf-bier-multicast-as-a-service-02, 4 January 2023, <https://www.ietf.org/archive/id/draft-ietf-bier-multicast-as-a-service-02.txt>.

[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, <https://www.rfc-editor.org/info/rfc7938>.

Authors' Addresses

Xiaohu Xu
Midea Group

Email: xuxh81@midea.com

Mach Chen
Huawei

Email: mach.chen@huawei.com

Keyur Patel
Arrcus, Inc.

Email: keyur@arrcus.com

IJsbrand Wijnands
Individual

Email: ice@braindump.be

Antoni Przygienda
Juniper

Email: prz@juniper.net

Zhaohui Zhang (editor)
Juniper

Email: zzhang@juniper.net