

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 6, 2021

H. Chen
Futurewei
A. Wang
China Telecom
G. Mishra
Verizon Inc.
Y. Fan
Casa Systems
L. Liu
Fujitsu
X. Liu
Volta Networks
November 2, 2020

BIER Egress Protection
draft-chen-bier-egress-protect-00

Abstract

This document describes a mechanism for fast protection against the failure of an egress node of a "Bit Index Explicit Replication" (BIER) domain. It does not have any per-flow state in the core of the domain. For a multicast packet to an egress node of the domain, when the egress node fails, its upstream hop as a PLR sends the packet to the egress' backup node once the PLR detects the failure.

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] [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 May 6, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Terminology	3
2. Overview of BIER Egress Protection	4
3. Protocol Extensions	5
3.1. Extensions to OSPF	5
3.2. Extensions to IS-IS	6
4. BIER Extensions	7
4.1. Egress Protection Bit Index Routing Tables	7
4.2. Egress Protection Bit Index Forwarding Tables	8
4.3. Updated Forwarding Procedure	8
4.4. Switching between EP and Normal Forwarding	9
5. Example Application of BIER Egress Protection	10
5.1. Example BIER Topology	10
5.2. BIRT and BIFT on a BFR	11
5.3. EP-BIRTs and EP-BIFTs on a BFR	12
5.4. Forwarding using EP-BIFT	14
6. Security Considerations	16
7. IANA Considerations	16
8. Acknowledgements	16
9. References	16
9.1. Normative References	16
9.2. Informative References	17
Authors' Addresses	18

1. Introduction

[RFC8279] specifies "Bit Index Explicit Replication" (BIER). It provides optimal forwarding of multicast data packets through a "multicast/BIER domain". It does not require the use of a protocol

for explicitly building multicast distribution trees, and it does not require intermediate nodes to maintain any per-flow state.

This document describes a mechanism for fast protection against the failure of an egress node of a "Bit Index Explicit Replication" (BIER) domain, which is called BIER Egress Protection.

This BIER Egress Protection does not require intermediate nodes to maintain any per-flow state for fast protection against the failure of an egress node of the flow.

1.1. Terminology

BFR: Bit-Forwarding Router.

BFIR: Bit-Forwarding Ingress Router.

BFER: Bit-Forwarding Egress Router.

BFR-id: BFR Identifier. It is a number in the range [1,65535].

BFR-NBR: BFR Neighbor.

F-BM: Forwarding Bit Mask.

BFR-prefix: An IP address (either IPv4 or IPv6) of a BFR.

BIRT: Bit Index Routing Table. It is a table that maps from the BFR-id (in a particular sub-domain) of a BFER to the BFR-prefix of that BFER, and to the BFR-NBR on the path to that BFER.

BIFT: Bit Index Forwarding Table.

FRR: Fast Re-Route.

PLR: Point of Local Repair.

LFA: Loop-Free Alternate.

RLFA: Remote LFA.

DLFA: Remote LFA with Directed forwarding.

IGP: Interior Gateway Protocol.

LSDB: Link State DataBase.

SPF: Shortest Path First.

SPT: Shortest Path Tree.

OSPF: Open Shortest Path First.

IS-IS: Intermediate System to Intermediate System.

LSA: Link State Advertisement in OSPF.

LSP: Link State Protocol Data Unit (PDU) in IS-IS.

SPT-old(R): The SPT rooted at node R using LSDB before X fails (i.e., old LSDB).

SPT-new(R, X): The SPT rooted at node R using LSDB without X after X fails (i.e., new LSDB).

P-Space $P(R, X)$: The set of nodes that are reachable from R without going through X. In other words, it is the set of nodes that are not downstream of X in SPT-old(R).

Extended P-Space $P'(R, X)$: The set of nodes that are reachable from R or a neighbor of R, without going through X.

Q-Space $Q(D, X)$: The set of nodes that do not use X to reach destination D using the old LSDB.

PQ node(R, X): A member of both the P-Space $P(R, X)$ (or the extended P-Space $P'(R, X)$) and the Q-Space $Q(D, X)$.

2. Overview of BIER Egress Protection

For fast protecting an egress node of a BIER domain, a backup egress node is configured on the egress node. After the configuration, the egress node distributes the information about the backup egress to its direct neighbors.

For clearly distinguishing between an egress node and a backup egress node, an egress node is called a primary egress node sometimes.

For a multicast packet to a primary egress node of the domain, when the primary egress node fails, its upstream hop as a point of local repair (PLR) sends the packet to the backup egress node configured to protect the primary egress node once the PLR detects the failure. The upstream hop of the primary egress node is its direct neighbor.

A Bit-Forwarding Router (BFR) in a BIER sub-domain builds and maintains an "Egress Protection Bit Index Routing Table" (EP-BIRT) for each of its BFR Neighbors (BFR-NBRs) that are egress nodes of the

domain to provide fast protection against the failure of an egress node. The BFR builds each EP-BIRT based on a BIRT defined in [RFC8279]. An "Egress Protection Bit Index Forwarding Table" (EP-BIFT) is derived from an EP-BIRT in a way that is similar to the way in which a BIFT is derived from a BIRT, which is defined in [RFC8279].

Once the BFR as a PLR detects the failure of its BFR-NBR X that is a primary egress node of the domain, for a multicast packet targeting to the primary egress node, the PLR uses the EP-BIFT for X to send the packet to the backup egress node configured to protect the primary egress node.

3. Protocol Extensions

This section defines extensions to OSPF and IS-IS for advertising the backup information (including the backup egress node for protecting a primary egress node) to its direct neighbors.

3.1. Extensions to OSPF

When a node P (as a primary egress node) has a backup egress node configured to protect against its failure, node P advertises the information about the backup egress node to its neighbors in its router information opaque LSA of LS type 9. The information is included in a backup egress node TLV. The format of the TLV is shown in Figure 1.

After each of the neighbors receives the backup egress node TLV from node P, it knows that node P as a primary egress node will be protected by the backup egress node in the TLV. Once detecting the failure of node P, it sends the packet targeting to node P towards the backup egress node.

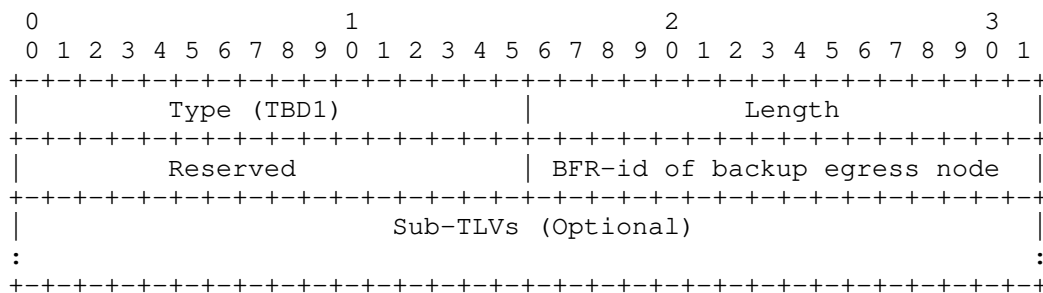


Figure 1: OSPF Backup Egress TLV

Type: 2 octets, its value (TBD1) is to be assigned by IANA.

Length: 2 octets, its value is 4 plus the length of the Sub-TLVs included. If no Sub-TLV is included, its value is 4.

Reserved: 2 octets, it MUST be set to zero when sending and be ignored while receiving.

BFR-id of backup egress node: 2 octets, its value is the BFR-id of the backup egress node configured to protect against the failure of the primary egress node (i.e., the node advertising this TLV).

Sub-TLVs (Optional): No Sub-TLV is defined now.

3.2. Extensions to IS-IS

For supporting fast protection against the failure of a primary egress node in a BIER domain, a new IS-IS TLV, called IS-IS backup egress node TLV, is defined. It contains the BFR-id of a backup egress node.

When a node P (as a primary egress node) has a backup egress node configured to protect against its failure, node P advertises the information about the backup egress node to its neighbors using a IS-IS backup egress node TLV.

This TLV may be advertised in IS-IS Hello (IIH) PDUs, LSPs, or in Circuit Scoped Link State PDUs (CS-LSP) [RFC7356]. The format of the TLV is shown in Figure 2.

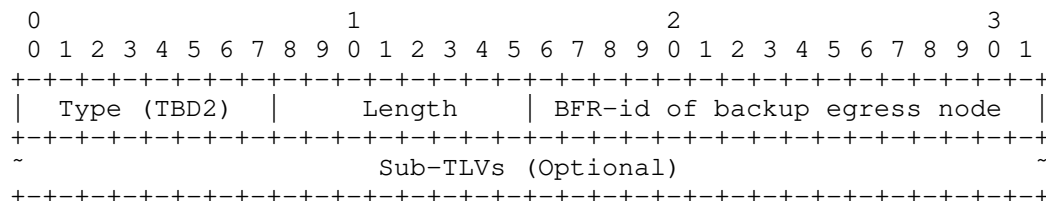


Figure 2: IS-IS Backup Egress TLV

Type: 1 octet, its value (TBD2) is to be assigned by IANA.

Length: 1 octet, its value is 2 plus the length of the Sub-TLVs included. If no Sub-TLV is included, its value is 2.

BFR-id of backup egress node: 2 octets, its value is the BFR-id of the backup egress node configured to protect against the failure of the primary egress node (i.e., the node advertising this TLV).

Sub-TLVs (Optional): No Sub-TLV is defined now.

4. BIER Extensions

4.1. Egress Protection Bit Index Routing Tables

If a BFR is a direct neighbor of an egress node in a BIER sub-domain, it builds and maintains a number of "Egress Protection Bit Index Routing Tables" (EP-BIRTs). There is an EP-BIRT for each of the BFR's neighbors that are egress nodes of the domain. The BFR builds each EP-BIRT based on its BIRT. Comparing to the BIRT, an EP-BIRT has a piece of new backup information for each BFER.

The new backup information for a BFER indicates if the BFER as an egress node is protected by the BFR. If so, the information further includes the backup egress node configured to protect the BFER.

In one implementation, the new backup information is represented by {EP, BE-BFER}. EP (short for Egress Protection) is a flag, indicating whether the BFER as an egress node is protected. EP = 1 means that the BFER is protected. EP = 0 means that the BFER is not protected. BE-BFER (short for Backup Egress BFER) is the BFER (i.e., BFER-id) of the backup egress node when EP = 1. BE-BFER is NULL (0) when EP = 0.

In the EP-BIRT for BFR-NBR X that is an egress node, the row having X as BFER and as its next hop BFR-NBR contains the new backup information {EP = 1, BE-BFER}, where BE-BFER is the BFER (i.e., BFER-id) of backup egress node for protecting the egress node. Each of the other rows in the EP-BIRT contains the new backup information {EP = 0, BE-BFER = NULL}.

When the egress node fails, for a multicast packet targeting to the primary egress node BFER (PE-BFER), the BFR sends the packet to the BE-BFER through using the route to the backup egress node. The BFR clears the bit for PE-BFER and adds the bit for BE-BFER in the packet's BitString first, and then forwards the packet according to the forwarding entry for BE-BFER.

The EP-BIRT for BFR-NBR X that is an egress node considers the failure of X. It has a route or say a next hop (i.e., BFR-NBR N on the path, where N is not X) to every BFER except for X.

The BFR may build the EP-BIRT for BFR-NBR X by copying its BIRT to the EP-BIRT and sets the new information for each BFER to empty such as {EP = 0, BE-BFER = NULL} first. And then it updates each of the rows in the EP-BIRT that has X as BFER or next hop BFR-NBR X.

For the BFR-id of a BFER in the EP-BIRT for egress node X, when the next hop BFR-NBR on the path to the BFER is X, the BFR checks whether

the BFER is X. If the BFER is not X, the BFR changes next hop BFR-NBR X to a backup next hop (BNH) when there is a BNH on a backup path to the BFER without going through X and the link from the BFR to X. If the BFER is X, the BFR adds the new backup information {EP = 1, BE-BFER} for the BFER as PE-BFER.

If there is not any BNH to a BFER to protect against the failure of X, the next hop BFR-NBR X to the BFER in the EP-BIRT for BFR-NBR X is changed to NULL. For a multicast packet having the BFER as one of its destinations, if the next hop BFR-NBR to the BFER is NULL, the BFR does not send the packet to the next hop BFR-NBR NULL but drops it when X fails.

Note: In another option, the next hop BFR-NBR X to the BFER in the EP-BIRT for BFR-NBR X keeps unchanged when there is not any BNH to the BFER to protect against the failure of X. In this case, for a multicast packet having the BFER as one of its destinations, the BFR sends the packet to X when X fails.

In one implementation, the BNH is the Loop-Free Node-Protecting Alternate defined in [RFC5286] to protect against the failure of X and link from the BFR to X. In another implementation, the BNH is the virtual Loop-Free Alternate (LFA), i.e., PQ node, defined in [RFC7490]. In a special case, a PQ node is a Loop-Free Node-Protecting Alternate defined in [RFC5286].

4.2. Egress Protection Bit Index Forwarding Tables

From each EP-BIRT on the BFR, an "Egress Protection Bit Index Forwarding Table" (EP-BIFT) is derived. In addition to having a route to a BFER in each row of the EP-BIFT which is the same as the EP-BIRT, it has a "Forwarding Bit Mask" (F-BM) in its each row. For the rows in the EP-BIRT that have the same SI and the same BFR-NBR and the same new backup information {EP, BE-BFER}, the F-BM for each of these rows in the EP-BIFT is the logical OR of the BitStrings of these rows.

This EP-BIFT is programmed into the data plane and is not used to forward any packet in normal operations. It is activated to forward a packet with a BIER header once the BFR detects the failure of BFR-NBR. The header contains SI, BitString, BitStringLength, and sub-domain.

4.3. Updated Forwarding Procedure

The forwarding procedure defined in [RFC8279] is updated/enhanced for an EP-BIFT to consider the egress protection (i.e., the new information {EP, BE-BFER} in the EP-BIFT). For a multicast packet

with the BitString indicating a BFER as one of its destinations, the updated forwarding procedure sends the packet towards the backup egress node of the BFER if the BFER is protected. It checks whether EP = 1 in the forwarding entry for the BFER. If EP = 1, the procedure clears the bit for the BFER as PE-BFER and adds the bit for BE-BFER in the packet's BitString first, and then forwards the packet using the row (i.e., forwarding entry) for BE-BFER.

The updated procedure is described in Figure 3. It is used with an EP-BIFT for BFR-NBR X as egress node on a BFR to forward multicast packets when X fails. It can also be used with a BIFT on the BFR to forward multicast packets in normal operations if the new backup information in each row of the BIFT is empty such as {EP = 0, BE-BFER = NULL}.

```

Packet = the packet received by BFR;
FOR each BFER k (from the rightmost in Packet's BitString) {
  IF BFER k is the BFR itself {
    copies Packet, sends the copy to the multicast
    flow overlay and clears bit k in Packet's BitString
  } ELSE {
    finds the row in the EP-BIFT for the sub-domain using
    Packet's SI and BitString as the key/index
    IF EP == 1 {
      clears bit k in Packet's BitString; //BFER k is PE-BFER
      adds bit j in Packet's BitString; //BFER j is BE-BFER
    } ELSE {
      IF BFR-NBR in the row is not NULL {
        Copies Packet, updates the copy's BitString by ANDing
        it with F-BM in the row, sends updated copy to BFR-NBR
      } // BFR-NBR == NULL, not sent Packet to BFR-NBR
      updates Packet's BitString by ANDing it with the INVERSE
      of the F-BM in the row
    }
  }
}

```

Figure 3: Updated Forwarding Procedure

4.4. Switching between EP and Normal Forwarding

The EP-BIFTs will be pre-computed and installed ready for activation when an egress node failure is detected. Once the BFR detects the failure of its BFR-NBR X as an egress, it activates the EP-BIFT for X to forward packets with BIER headers and de-activates its BIFT. After activation of the EP-BIFT, it remains in effect until it is no longer required.

In general, when the routing protocol has re-converged on the new topology taking into account the failure of X, the BIRT is re-computed using the updated LSDB and the BIFT is re-derived from the BIRT. Once the BIFT is installed ready for activation, it is activated to forward packets with BIER headers and the EP-BIFT for X is de-activated.

From the new topology, the BFR computes/re-computes the EP-BIRT for each BFR-NBR Y as an egress of the BFR and the EP-BIFT for Y is derived/re-derived from the EP-BIRT for Y. The EP-BIFT is installed/re-installed ready for activation when Y fails.

5. Example Application of BIER Egress Protection

This section illustrates an example application of BIER Egress Protection on a BFR in a BIER topology in Figure 4.

5.1. Example BIER Topology

An example BIER topology for a BIER sub-domain is shown in Figure 4. It has 8 nodes/BFRs A, B, C, D, E, F, G and H. Each of the links connecting these nodes/BFRs has a cost. The link cost of 1 is default and is not indicated in the figure. The link costs of other values such as 2 and 3 are indicated in the figure.

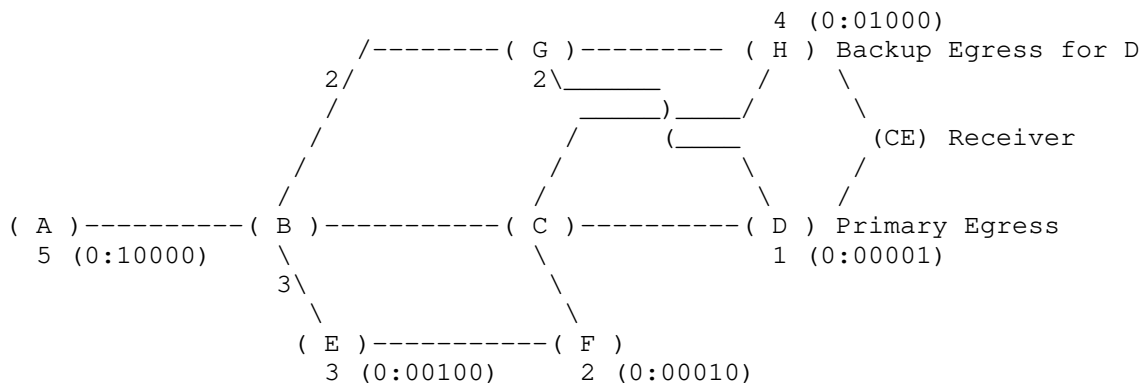


Figure 4: Example BIER Topology

Nodes/BFRs D, F, E, H and A are BFRs and have BFR-ids 1, 2, 3, 4, and 5 respectively. For simplicity, these BFR-ids are represented by (SI:BitString), where SI = 0 and BitString is of 5 bits. BFR-ids 1, 2, 3, 4, and 5 are represented by (0:00001), (0:00010), (0:00100), (0:01000) and (0:10000) respectively.

BFER H is configured to protect BFER D on BFR D. Suppose that this information is distributed to BFR D's neighbors BFR C and BFR G by IGP. BFR C and BFR G know that H is the backup egress to protect the primary egress D.

CE is a multicast traffic Receiver, which is dual homed to primary egress node D and backup egress node H for protecting primary egress D. During normal operations, there is no multicast traffic to CE from backup egress node H and CE receives the multicast traffic only from primary egress node D. There is no duplicated traffic to receiver CE. This is different from MoFRR in [RFC7431], where duplicated traffic is sent to a dual homed receiver. When primary egress node D fails, the multicast traffic is sent to CE from backup egress node H.

5.2. BIRT and BIFT on a BFR

Every BFR in a BIER sub-domain/topology builds and maintains a Bit Index Routing Table (BIRT). For the BIER topology in Figure 4, each of 8 nodes/BFRs A, B, C, D, E, F, G and H builds and maintains a BIRT using the LSDB for the topology.

The BIRT built on BFR C (i.e. node C) is shown in Figure 5.

BFR-id (SI:BitString)	BFR-Prefix of Dest BFER	BFR-NBR (Next Hop)
1 (0:00001)	D	D
2 (0:00010)	F	F
3 (0:00100)	E	F
4 (0:01000)	H	H
5 (0:10000)	A	B

Figure 5: Bit Index Routing Table on BFR C

The 1st row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER D with BFR-id 1 is BFR D.

The 2nd row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER F with BFR-id 2 is BFR F.

The 3rd row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER E with BFR-id 3 is BFR F.

The 4-th row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER H with BFR-id 4 is BFR H.

The 5-th row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER A with BFR-id 5 is BFR B.

From this BIRT on BFR C, a Bit Index Forwarding Table (BIFT) is derived. This BIFT is shown in Figure 6.

The 2nd and 3-th rows in the BIRT have the same SI = 0 and next hop BFR-NBR = F. The F-BM for each of these two rows in the BIFT is the logical OR of the BitStrings of these rows, which is 00110 (00010 OR 00100 = 00110).

The F-BM for 1st row in the BIFT is 00001.

The F-BM for 4-th row in the BIFT is 01000.

The F-BM for 5-th row in the BIFT is 10000.

BFR-id (SI:BitString)	F-BM	BFR-NBR (Next Hop)
1 (0:00001)	00001	D
2 (0:00010)	00110	F
3 (0:00100)	00110	F
4 (0:01000)	01000	H
5 (0:10000)	10000	B

Figure 6: Bit Index Forwarding Table on BFR C

5.3. EP-BIRTs and EP-BIFTs on a BFR

Each of the BFRs that are neighbors of egress nodes (i.e., BFERs) in a BIER sub-domain/topology builds and maintains a number of Egress Protection Bit Index Routing Tables (EP-BIRTs).

For the BIER topology in Figure 4,

BFR B is the neighbor of BFERs A and E;
 BFR C is the neighbor of BFERs D, F and H;
 BFR E is the neighbor of BFER F;
 BFR F is the neighbor of BFER E;
 BFR G is the neighbor of BFERs D and H.

Each of 5 nodes/BFRs B, C, E, F and G builds and maintains a number of EP-BIRTs using the LSDB for the topology for its every BFR-NBR as an egress node.

For example, BFR C (i.e., node C) in the BIER topology builds and maintains three EP-BIRTs for its three BFR-NBRs (BFERs D, F and H) that are egress nodes respectively.

The EP-BIRT for BFER D built by BFR C based on the BIRT on BFR C (refer to Figure 5) is shown in Figure 7.

The BIRT is copied to the EP-BIRT for BFER D (i.e., the first three columns of the EP-BIRT). The new backup information (i.e., the 4-th column) for every row in the EP-BIRT is initialized to {EP = 0, BE-BFER = 0/NULL}.

BFR-id (SI:BitString)	BFR-Prefix of Dest BFER	BFR-NBR (Next Hop)	{EP, BE-BFER} (Backup Info)
1 (0:00001)	D	D/NULL	EP=1, BE-BFER=H
2 (0:00010)	F	F	EP=0, BE-BFER=0
3 (0:00100)	E	F	EP=0, BE-BFER=0
4 (0:01000)	H	H	EP=0, BE-BFER=0
5 (0:10000)	A	B	EP=0, BE-BFER=0

Figure 7: EP-BIRT for BFER D on BFR C

In the EP-BIRT for BFER D, the row that has BFR-NBR == D is the 1st row. This row has the new backup information {EP = 1, BE-BFER = H}, which indicates that BFER D (i.e., primary egress node D) is protected by BFER H (i.e., backup egress node H). Each of the other rows has the new backup information {EP = 0, BE-BFER = 0/NULL}.

The 1st row in the EP-BIRT indicates that the next hop BFR-NBR on the path to BFER D with BFR-id 1 is NULL (changed to NULL from D). There is no backup next hop (BNH) to D when D fails.

The 2nd row in the EP-BIRT indicates that the next hop BFR-NBR on the path to BFER F with BFR-id 2 is BFR F.

The 3rd row in the EP-BIRT indicates that the next hop BFR-NBR on the path to BFER E with BFR-id 3 is BFR F.

The 4-th row in the EP-BIRT indicates that the next hop BFR-NBR on the path to BFER H with BFR-id 4 is BFR H.

The 5-th row in the EP-BIRT indicates that the next hop BFR-NBR on the path to BFER A with BFR-id 5 is BFR B.

From this EP-BIRT for BFER D on BFR C, an Egress Protection Bit Index Forwarding Table (EP-BIFT) is derived. This EP-BIFT for BFER D is shown in Figure 8.

The 2nd and 3rd rows in the EP-BIRT have the same SI = 0, the same next hop BFR-NBR = E and the same backup information {EP=0, BE-BFER=0}. The F-BM for each of these two rows in the EP-BIFT is the logical OR of the BitStrings of these rows, which is 00110 (00010 OR 00100 = 00110).

BFR-id (SI:BitString)	F-BM	BFR-NBR (Next Hop)	{EP, BE-BFER} (Backup Info)
1 (0:00001)	00001	NULL	EP=1, BE-BFER=H
2 (0:00010)	00110	F	EP=0, BE-BFER=0
3 (0:00100)	00110	F	EP=0, BE-BFER=0
4 (0:01000)	01000	H	EP=0, BE-BFER=0
5 (0:10000)	10000	B	EP=0, BE-BFER=0

Figure 8: EP-BIFT for BFER D on BFR C

The F-BM for 1st row in the EP-BIFT is 00001.

The F-BM for 4-th row in the EP-BIFT is 01000.

The F-BM for 5-th row in the EP-BIFT is 10000.

5.4. Forwarding using EP-BIFT

Suppose that there is a multicast traffic from BFR A as ingress/BFIR to egresses/BFERs D, F and E. For every packet of the traffic, after receiving it, BFR A adds a BIER header into the packet and sends the

packet with the BIER header to BFR B, which sends the packet BFR C. The BIER header contains (SI:BitString) = (0:00111) for egresses/BFRs D, F and E.

In normal operations, after receiving the packet from BFR B, BFR C copies, updates and sends the packet to BFR D and BFR F using the BIFT on BFR C according to the forwarding procedure defined in [RFC8279].

Once BFR C detects the failure of its BFR-NBR D, which is a BFER, after receiving the packet from BFR B, BFR C copies, updates and sends the packet using the EP-BIFT for BFER D on BFR C according to the updated forwarding procedure.

For the packet targeting to BFER D (i.e., primary egress node), BFR C sends it towards BFER H (i.e., backup egress node), which is configured to protect BFER D.

For example, once BFR C detects the failure of its BFR-NBR D, after receiving the packet from BFR B, BFR C copies, updates and sends the packet to BFR H and BFR F using the EP-BIFT for BFER D on BFR C.

The packet received by BFR C from BFR B contains (SI:BitString) = (0:00111). The rightmost one bit in BitString is bit 1. For BFER 1 (0:00001) (i.e., BFR D as BFER), BFR C gets the 1st row (i.e., forwarding entry) in the EP-BIFT for BFER D. EP = 1 in the row indicates that BFER D is protected against the failure of D. BFR C clears bit 1 in Packet's BitString and sets bit 4 (i.e., the bit for BE-BFER = H) in Packet's BitString to one. The BitString in Packet is 01110 now. This lets BFR C send Packet to BE-BFER H.

For the packet containing BitString = 01110, the rightmost one bit in BitString is bit 2. For BFER 2 (0:00010) (i.e., BFR F as BFER), BFR C gets the 2nd row (i.e., forwarding entry) in the EP-BIFT for BFER D. EP = 0 and the next hop BFR-NBR is F in the row. BFR C copies, updates and sends the packet to F.

The packet sent to F contains the updated BitString = 00110, which is 01110 & F-BM in the 2nd row = 01110 & 00110 = 00110.

After sending the packet to F, BFR C updates the original packet by ANDing its BitString with the INVERSE of the F-BM in the 2nd row. The updated BitString = 01000, which is 01110 & ~F-BM in the row = 01110 & 11001 = 01000.

For the packet containing BitString = 01000, the rightmost one bit in BitString is bit 4. For BFER 4 (0:01000) (i.e., BFR H as BFER), BFR C gets the 4-th row (i.e., forwarding entry) in the EP-BIFT for BFER

D. EP = 0 and the next hop BFR-NBR is H in the row. BFR C copies, updates and sends the packet to H. The packet sent to H contains BitString = 01000.

After sending the packet to H, BFR C updates the original packet by ANDing its BitString with the INVERSE of the F-BM in the 4-th row. The updated BitString = 00000, which is 01000 & ~F-BM in the row = 01000 & 10111 = 00000.

The updated packet has BitString without any one bit. BFR C finishes forwarding the packet to F and H (backup for D). BFR F will send the packet to E.

6. Security Considerations

TBD.

7. IANA Considerations

No requirements for IANA.

8. Acknowledgements

The authors would like to thank people for their comments to this work.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, DOI 10.17487/RFC5226, May 2008, <<https://www.rfc-editor.org/info/rfc5226>>.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, DOI 10.17487/RFC5250, July 2008, <<https://www.rfc-editor.org/info/rfc5250>>.
- [RFC5286] Atlas, A., Ed. and A. Zinin, Ed., "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, DOI 10.17487/RFC5286, September 2008, <<https://www.rfc-editor.org/info/rfc5286>>.

- [RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, DOI 10.17487/RFC5714, January 2010, <<https://www.rfc-editor.org/info/rfc5714>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, DOI 10.17487/RFC7356, September 2014, <<https://www.rfc-editor.org/info/rfc7356>>.
- [RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <<https://www.rfc-editor.org/info/rfc7490>>.
- [RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, DOI 10.17487/RFC7684, November 2015, <<https://www.rfc-editor.org/info/rfc7684>>.
- [RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 7770, DOI 10.17487/RFC7770, February 2016, <<https://www.rfc-editor.org/info/rfc7770>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [RFC8556] Rosen, E., Ed., Sivakumar, M., Przygienda, T., Aldrin, S., and A. Dolganow, "Multicast VPN Using Bit Index Explicit Replication (BIER)", RFC 8556, DOI 10.17487/RFC8556, April 2019, <<https://www.rfc-editor.org/info/rfc8556>>.

9.2. Informative References

- [I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B.,
Francois, P., Voyer, D., Clad, F., and P. Camarillo,
"Topology Independent Fast Reroute using Segment Routing",
draft-ietf-rtgwg-segment-routing-ti-lfa-04 (work in
progress), August 2020.
- [I-D.ietf-spring-segment-protection-sr-te-paths]
Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu,
"Segment Protection for SR-TE Paths", draft-ietf-spring-
segment-protection-sr-te-paths-00 (work in progress),
September 2020.
- [RFC7431] Karan, A., Filsfils, C., Wijnands, IJ., Ed., and B.
Decraene, "Multicast-Only Fast Reroute", RFC 7431,
DOI 10.17487/RFC7431, August 2015,
<<https://www.rfc-editor.org/info/rfc7431>>.
- [RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A.,
Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation
for Bit Index Explicit Replication (BIER) in MPLS and Non-
MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January
2018, <<https://www.rfc-editor.org/info/rfc8296>>.
- [RFC8401] Ginsberg, L., Ed., Przygienda, T., Aldrin, S., and Z.
Zhang, "Bit Index Explicit Replication (BIER) Support via
IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018,
<<https://www.rfc-editor.org/info/rfc8401>>.
- [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>>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA
USA

Email: Huaimo.chen@futurewei.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing, 102209
China

Email: wangaj3@chinatelecom.cn

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring MD 20904
USA

Phone: 301 502-1347
Email: gyan.s.mishra@verizon.com

Yanhe Fan
Casa Systems
USA

Email: yfan@casa-systems.com

Lei Liu
Fujitsu

USA

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks

McLean, VA
USA

Email: xufeng.liu.ietf@gmail.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 6, 2021

H. Chen
Futurewei
A. Wang
China Telecom
G. Mishra
Verizon Inc.
Y. Fan
Casa Systems
L. Liu
Fujitsu
X. Liu
Volta Networks
November 2, 2020

BIER Fast ReRoute
draft-chen-bier-frr-00

Abstract

This document describes a mechanism for fast re-route (FRR) protection against the failure of a node or link in the core of a "Bit Index Explicit Replication" (BIER) domain. It does not have any per-flow state in the core. For a multicast packet to traverse a node in the domain, when the node fails, its upstream hop as a PLR reroutes the packet around the failed node once it detects the failure.

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] [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 May 6, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Terminology	3
2. BIER FRR Solution	4
2.1. Overview of BIER forwarding	4
2.2. FRR Bit Index Routing Tables	5
2.3. FRR Bit Index Forwarding Tables	6
2.4. Updated Forwarding Procedure	7
2.5. Switching between FRR and Normal Forwarding	7
3. Example Application of BIER FRR	8
3.1. Example BIER Topology	8
3.2. BIRT and BIFT on a BFR	8
3.3. FRR-BIRTs and FRR-BIFTs on a BFR	10
3.4. Forwarding using FRR-BIFT	12
4. Security Considerations	13
5. IANA Considerations	13
6. Acknowledgements	13
7. References	13
7.1. Normative References	13
7.2. Informative References	15
Authors' Addresses	15

1. Introduction

[RFC8279] specifies "Bit Index Explicit Replication" (BIER). It provides optimal forwarding of multicast data packets through a "multicast/BIER domain". It does not require the use of a protocol

for explicitly building multicast distribution trees, and it does not require intermediate nodes to maintain any per-flow state.

This document describes a mechanism for fast re-route (FRR) protection against the failure of a node or link in the core of a BIER domain, which is called BIER-FRR. For a multicast packet with a BIER header to traverse a node in the domain, when the node fails, its upstream hop as a point of local repair (PLR) reroutes the packet around the failed node once it detects the failure.

This BIER-FRR does not require intermediate nodes to maintain any per-flow state for FRR protection against the failure of a node or link along the flow.

1.1. Terminology

BFR: Bit-Forwarding Router.

BFIR: Bit-Forwarding Ingress Router.

BFER: Bit-Forwarding Egress Router.

BFR-id: BFR Identifier. It is a number in the range [1,65535].

BFR-NBR: BFR Neighbor.

F-BM: Forwarding Bit Mask.

BFR-prefix: An IP address (either IPv4 or IPv6) of a BFR.

BIRT: Bit Index Routing Table. It is a table that maps from the BFR-id (in a particular sub-domain) of a BFER to the BFR-prefix of that BFER, and to the BFR-NBR on the path to that BFER.

BIFT: Bit Index Forwarding Table.

FRR: Fast Re-Route.

PLR: Point of Local Repair.

LFA: Loop-Free Alternate.

RLFA: Remote LFA.

DLFA: Remote LFA with Directed forwarding.

IGP: Interior Gateway Protocol.

LSDB: Link State DataBase.

SPF: Shortest Path First.

SPT: Shortest Path Tree.

SPT-old(R): The SPT rooted at node R using LSDB before X fails (i.e., old LSDB).

SPT-new(R, X): The SPT rooted at node R using LSDB without X after X fails (i.e., new LSDB).

P-Space $P(R, X)$: The set of nodes that are reachable from R without going through X. In other words, it is the set of nodes that are not downstream of X in SPT-old(R).

Extended P-Space $P'(R, X)$: The set of nodes that are reachable from R or a neighbor of R, without going through X.

Q-Space $Q(D, X)$: The set of nodes that do not use X to reach destination D using the old LSDB.

PQ node(R, X): A member of both the P-Space $P(R, X)$ (or the extended P-Space $P'(R, X)$) and the Q-Space $Q(D, X)$.

2. BIER FRR Solution

A Bit-Forwarding Router (BFR) in a BIER sub-domain builds and maintains a "FRR Bit Index Routing Table" (FRR-BIRT) for each of its BFR Neighbors (BFR-NBRs) to provide BIER-FRR. The BFR builds each FRR-BIRT based on a BIRT defined in [RFC8279]. A "FRR Bit Index Forwarding Table" (FRR-BIFT) is derived from a FRR-BIRT in the same way as a BIFT is derived from a BIRT, which is defined in [RFC8279].

The forwarding procedure defined in [RFC8279] is enhanced/updated for FRR-BIFTs. Once the BFR as a PLR detects the failure of its BFR-NBR X, it uses the FRR-BIFT for X to forward packets with BIER headers to get around failed X according to the updated/enhanced forwarding procedure.

2.1. Overview of BIER forwarding

This section briefs the BIRT, BIFT and forwarding procedure defined in [RFC8279].

There is a "Bit Index Routing Table" (BIRT) for a BIER sub-domain on a BFR. The BIRT maps the BFR Identifier (BFR-id) (in the sub-domain) of a Bit-Forwarding Egress Router (BFER) to the BFR-prefix of that

BFER, and to the BFR-NBR on the shortest path to that BFER. In other words, the BIRT has a route or say a next hop (i.e., BFR-NBR on the path) to every BFER.

From the BIRT on the BFR, a "Bit Index Forwarding Table" (BIFT) is derived. In addition to having a route to a BFER in each row of the BIFT which is the same as the BIRT, it has a "Forwarding Bit Mask" (F-BM) in its each row. For the rows in the BIRT that have the same SI and the same BFR-NBR, the F-BM for each of these rows in the BIFT is the logical OR of the BitStrings of these rows.

This BIFT is programmed into the data plane and used to forward a packet with a BIER header. The header contains SI, BitString, BitStringLength, and sub-domain.

When a BFR receives a packet, for each BFER k (from the rightmost to the leftmost) represented in the SI and BitString of the packet, if BFER k is the BFR itself, the BFR copies the packet, sends the copy to the multicast flow overlay and clears bit k in the original packet; otherwise the BFR finds the row (i.e., forwarding entry) in the BIFT for the sub-domain using the SI and BitString as the key or say index, and then copies, updates and forwards the packet to the BFR-NBR (i.e., the next hop) indicated by the row (i.e., forwarding entry).

After copying the packet and before forwarding it to the BFR-NBR, the packet's BitString is updated by ANDing it with the F-BM in the forwarding entry (i.e., $\text{PacketCopy} \rightarrow \text{BitString} \&= \text{F-BM}$).

After forwarding the updated packet to a BFR-NBR and before forwarding the original packet to another BFR-NBR, the original packet's BitString is changed by ANDing it with the INVERSE of the F-BM (i.e., $\text{Packet} \rightarrow \text{BitString} \&= \sim \text{F-BM}$).

2.2. FRR Bit Index Routing Tables

Each BFR in a BIER sub-domain builds and maintains a number of "FRR Bit Index Routing Tables" (FRR-BIRTs). There is a FRR-BIRT for each BFR-NBR of the BFR. The BFR builds each FRR-BIRT based on its BIRT. It has the same format as the BIRT.

The FRR-BIRT for BFR-NBR X of the BFR considers the failure of X and maps the BFR-id (in the sub-domain) of a BFER to the BFR-prefix of that BFER, and to BFR-NBR N on the path to that BFER. In other words, the FRR-BIRT has a route or say a next hop (i.e., BFR-NBR N on the path, where N is not X) to every BFER when BFR-NBR X fails.

The BFR may build the FRR-BIRT for BFR-NBR X by copying its BIRT to the FRR-BIRT first, and then change the next hop with value BFR-NBR X in the FRR-BIRT to a backup next hop (BNH) to protect against the failure of X. In other words, for the BFR-id of a BFER in the FRR-BIRT for BFR-NBR X, if the next hop BFR-NBR on the path to the BFER is X, it is changed to a BNH when there is a BNH on a backup path to the BFER without going through X and the link from the BFR to X.

If there is not any BNH to a BFER to protect against the failure of X, the next hop BFR-NBR X to the BFER in the FRR-BIRT for BFR-NBR X is changed to NULL. For a multicast packet having the BFER as one of its destinations, if the next hop BFR-NBR to the BFER is NULL, the BFR does not send the packet to the next hop BFR-NBR NULL but drops it when X fails.

Note: In another option, the next hop BFR-NBR X to the BFER in the FRR-BIRT for BFR-NBR X keeps unchanged when there is not any BNH to the BFER to protect against the failure of X. In this case, for a multicast packet having the BFER as one of its destinations, the BFR sends the packet to X when X fails.

In one implementation, the BNH is the Loop-Free Node-Protecting Alternate defined in [RFC5286] to protect against the failure of X and link from the BFR to X. In another implementation, the BNH is the virtual Loop-Free Alternate (LFA), i.e., PQ node, defined in [RFC7490]. In a special case, a PQ node is a Loop-Free Node-Protecting Alternate defined in [RFC5286].

2.3. FRR Bit Index Forwarding Tables

From each FRR-BIRT on the BFR, a "FRR Bit Index Forwarding Table" (FRR-BIFT) is derived. In addition to having a route to a BFER in each row of the FRR-BIFT which is the same as the FRR-BIRT, it has a "Forwarding Bit Mask" (F-BM) in its each row. For the rows in the FRR-BIRT that have the same SI and the same BFR-NBR, the F-BM for each of these rows in the FRR-BIFT is the logical OR of the BitStrings of these rows.

This FRR-BIFT is programmed into the data plane and is not used to forward any packet in normal operations. It is activated to forward a packet with a BIER header once the BFR detects the failure of BFR-NBR. The header contains SI, BitString, BitStringLength, and sub-domain.

2.4. Updated Forwarding Procedure

The forwarding procedure defined in [RFC8279] is updated/enhanced for a FRR-BIFT to consider the case where the next hop BFR-NBR to a BFER is NULL. For a multicast packet with the BitString indicating a BFER as one of its destinations, the updated forwarding procedure checks whether the next hop BFR-NBR to the BFER in the FRR-BIFT is NULL. If it is NULL, the procedure will not send the packet to this next hop BFR-NBR NULL but drop the packet.

The updated procedure is described in Figure 1. It is used with a FRR-BIFT for BFR-NBR X on a BFR to forward multicast packets when X fails. It can also be used with a BIFT on the BFR to forward multicast packets in normal operations.

```

Packet = the packet received by BFR;
FOR each BFER k (from the rightmost in Packet's BitString) {
  IF BFER k is the BFR itself {
    copies Packet, sends the copy to the multicast
    flow overlay and clears bit k in Packet's BitString
  } else {
    finds the row in the FRR-BIFT for the sub-domain using
    Packet's SI and BitString as the key/index
    IF BFR-NBR in the row is not NULL {
      Copies Packet, updates the copy's BitString by ANDing
      it with F-BM in the row, sends updated copy to BFR-NBR
    } // BFR-NBR == NULL, not sent Packet to BFR-NBR
    updates Packet's BitString by ANDing it with the INVERSE
    of the F-BM in the row
  }
}

```

Figure 1: Updated Forwarding Procedure

2.5. Switching between FRR and Normal Forwarding

The FRR-BIFTs will be pre-computed and installed ready for activation when a failure is detected. Once the BFR detects the failure of its BFR-NBR X, it activates the FRR-BIFT for X to forward packets with BIER headers and de-activates its BIFT. After activation of the FRR-BIFT, it remains in effect until it is no longer required.

In general, when the routing protocol has re-converged on the new topology taking into account the failure of X, the BIRT is re-computed using the updated LSDB and the BIFT is re-derived from the BIRT. Once the BIFT is installed ready for activation, it is activated to forward packets with BIER headers and the FRR-BIFT for X is de-activated.

From the new topology, the BFR computes/re-computes the FRR-BIRT for each BFR-NBR Y of the BFR and the FRR-BIFT for Y is derived/re-derived from the FRR-BIRT for Y. The FRR-BIFT is installed/re-installed ready for activation when Y fails.

3. Example Application of BIER FRR

This section illustrates an example application of BIER FRR on a BFR in a BIER topology in Figure 2.

3.1. Example BIER Topology

An example BIER topology for a BIER sub-domain is shown in Figure 2. It has 8 nodes/BFRs A, B, C, D, E, F, G and H. Each of the links connecting these nodes/BFRs has a cost. The link cost of 1 is default and is not indicated in the figure. The link cost of other value such as 2 is indicated in the figure.

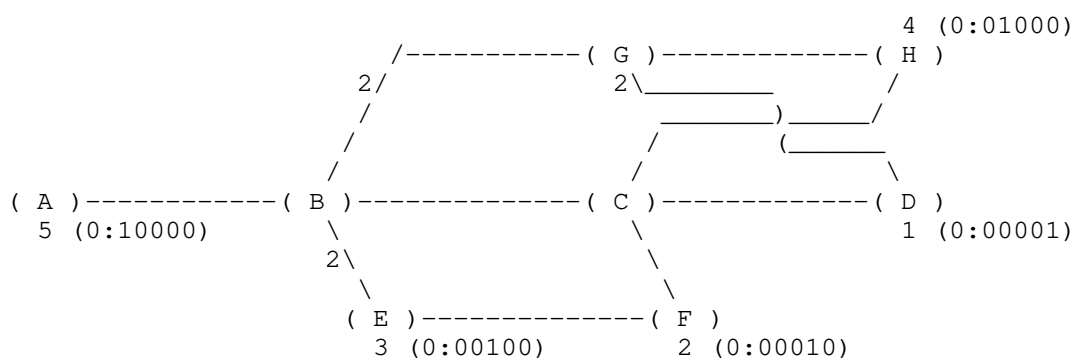


Figure 2: Example BIER Topology

Nodes/BFRs D, F, E, H and A are BFRs and have BFR-ids 1, 2, 3, 4, and 5 respectively. For simplicity, these BFR-ids are represented by (SI:BitString), where SI = 0 and BitString is of 5 bits. BFR-ids 1, 2, 3, 4, and 5 are represented by (0:00001), (0:00010), (0:00100), (0:01000) and (0:10000) respectively.

3.2. BIRT and BIFT on a BFR

Every BFR in a BIER sub-domain/topology builds and maintains a Bit Index Routing Table (BIRT). For the BIER topology in Figure 2, each of 8 nodes/BFRs A, B, C, D, E, F, G and H builds and maintains a BIRT using the LSDB for the topology.

The BIRT built on BFR B (i.e. node B) is shown in Figure 3.

BFR-id (SI:BitString)	BFR-Prefix of Dest BFER	BFR-NBR (Next Hop)
1 (0:00001)	D	C
2 (0:00010)	F	C
3 (0:00100)	E	E
4 (0:01000)	H	C
5 (0:10000)	A	A

Figure 3: Bit Index Routing Table on BFR B

The 1st row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER D with BFR-id 1 is BFR C.

The 2nd row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER F with BFR-id 2 is BFR C.

The 3rd row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER E with BFR-id 3 is BFR E.

The 4-th row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER H with BFR-id 4 is BFR C.

The 5-th row in the BIRT indicates that the next hop BFR-NBR on the shortest path to BFER A with BFR-id 5 is BFR A.

From this BIRT on BFR B, a Bit Index Forwarding Table (BIFT) is derived. This BIFT is shown in Figure 4.

The 1st, 2nd and 4-th rows in the BIRT have the same SI = 0 and next hop BFR-NBR = C. The F-BM for each of these three rows in the BIFT is the logical OR of the BitStrings of these rows, which is 01011 (00001 OR 00010 OR 01000 = 01011).

The F-BM for 3rd row in the BIFT is 00100. The F-BM for 5-th row in the BIFT is 10000.

BFR-id (SI:BitString)	F-BM	BFR-NBR (Next Hop)
1 (0:00001)	01011	C
2 (0:00010)	01011	C
3 (0:00100)	00100	E
4 (0:01000)	01011	C
5 (0:10000)	10000	A

Figure 4: Bit Index Forwarding Table on BFR B

3.3. FRR-BIRTs and FRR-BIFTs on a BFR

Every BFR in a BIER sub-domain/topology builds and maintains a number of FRR Bit Index Routing Tables (FRR-BIRTs). For the BIER topology in Figure 2, each of 8 nodes/BFRs A, B, C, D, E, F, G and H builds and maintains a number of FRR-BIRTs using the LSDB for the topology for its every BFR-NBR.

For example, BFR B (i.e., node B) in the BIER topology builds and maintains four FRR-BIRTs for its four BFR-NBRs (BFR C, BFR E, BFR A and BFR G) respectively. The FRR-BIRT for BFR C built by BFR B is shown in Figure 5.

BFR-id (SI:BitString)	BFR-Prefix of Dest BFER	BFR-NBR (Next Hop)
1 (0:00001)	D	G
2 (0:00010)	F	E
3 (0:00100)	E	E
4 (0:01000)	H	G
5 (0:10000)	A	A

Figure 5: FRR Bit Index Routing Table for BFR C on BFR B

The 1st row in the FRR-BIRT indicates that the next hop BFR-NBR on the path to BFER D with BFR-id 1 is BFR G. G is the Loop-Free Node-Protecting Alternate defined in [RFC5286] to protect against the failure of C and link from B to C.

The 2nd row in the FRR-BIRT indicates that the next hop BFR-NBR on the path to BFER F with BFR-id 2 is BFR E. E is the Loop-Free Node-Protecting Alternate defined in [RFC5286] to protect against the failure of C and link from B to C.

The 3rd row in the FRR-BIRT indicates that the next hop BFR-NBR on the path to BFER E with BFR-id 3 is BFR E.

The 4-th row in the FRR-BIRT indicates that the next hop BFR-NBR on the path to BFER H with BFR-id 4 is BFR G. G is the Loop-Free Node-Protecting Alternate defined in [RFC5286] to protect against the failure of C and link from B to C.

The 5-th row in the FRR-BIRT indicates that the next hop BFR-NBR on the path to BFER A with BFR-id 5 is BFR A.

From this FRR-BIRT for BFR C on BFR B, a FRR Bit Index Forwarding Table (FRR-BIFT) is derived. This FRR-BIFT for BFR C is shown in Figure 6.

The 1st and 4-th rows in the FRR-BIRT have the same SI = 0 and next hop BFR-NBR = G. The F-BM for each of these two rows in the FRR-BIFT is the logical OR of the BitStrings of these rows, which is 01001 (00001 OR 01000 = 01001).

BFR-id (SI:BitString)	F-BM	BFR-NBR (Next Hop)
1 (0:00001)	01001	G
2 (0:00010)	00110	E
3 (0:00100)	00110	E
4 (0:01000)	01001	G
5 (0:10000)	10000	A

Figure 6: FRR Bit Index Forwarding Table for BFR C on BFR B

The 2nd and 3rd rows in the FRR-BIFT have the same SI = 0 and next hop BFR-NBR = E. The F-BM for each of these two rows in the FRR-BIFT is the logical OR of the BitStrings of these rows, which is 00110 (00010 OR 00100 = 00110).

The F-BM for 5-th row in the FRR-BIFT is 10000.

3.4. Forwarding using FRR-BIFT

Suppose that there is a multicast traffic from BFR A as ingress/BFIR to egresses/BFERs D, F, E and H. For every packet of the traffic, after receiving it, BFR A adds a BIER header into the packet and sends the packet with the BIER header to BFR B. The BIER header contains (SI:BitString) = (0:01111) for egresses/BFERs D, F, E and H.

In normal operations, after receiving the packet from BFR A, BFR B copies, updates and sends the packet to BFR C and BFR E using the BIFT on BFR B according to the forwarding procedure defined in [RFC8279].

Once BFR B detects the failure of its BFR-NBR X, after receiving the packet from BFR A, BFR B copies, updates and sends the packet using the FRR-BIFT for X on BFR B to avoid X and link from B to X according to the forwarding procedure defined in [RFC8279].

For example, once BFR B detects the failure of its BFR-NBR C, after receiving the packet from BFR A, BFR B copies, updates and sends the packet to BFR G and BFR E using the FRR-BIFT for BFR C on BFR B to avoid C and link from B to C.

The packet received by BFR B from BFR A contains (SI:BitString) = (0:01111). The rightmost one bit in BitString is bit 1. For BFER 1 (0:00001) (i.e., BFR D as BFER), BFR B gets the 1st row (i.e., forwarding entry) in the FRR-BIFT for BFR C. The next hop BFR-NBR is G in the row. BFR B copies, updates and forwards the packet to G.

The packet sent to G contains the updated BitString = 01001, which is 01111 & F-BM in the row = 01111 & 01001.

After sending the packet to G, BFR B updates the original packet by ANDing its BitString with the INVERSE of the F-BM in the row. The updated BitString = 00110, which is 01111 & ~F-BM in the row = 01111 & 00110.

For the packet containing BitString = 00110, the rightmost one bit in BitString is bit 2. For BFER 2 (0:00010) (i.e., BFR F as BFER), BFR B gets the 2nd row (i.e., forwarding entry) in the FRR-BIFT for BFR

C. The next hop BFR-NBR is E in the row. BFR B copies, updates and forwards the packet to E.

The packet sent to E contains the updated BitString = 00110, which is 00110 & F-BM in the 2nd row = 00110 & 00110.

After sending the packet to E, BFR B updates the original packet by ANDing its BitString with the INVERSE of the F-BM in the 2nd row. The updated BitString = 00000, which is 00110 & ~F-BM in the row = 00110 & 11001.

The updated packet has BitString without any one bit. BFR B finishes forwarding the packet from A to D, F, E and H.

4. Security Considerations

TBD.

5. IANA Considerations

No requirements for IANA.

6. Acknowledgements

The authors would like to thank people for their comments to this work.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, DOI 10.17487/RFC5226, May 2008, <<https://www.rfc-editor.org/info/rfc5226>>.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, DOI 10.17487/RFC5250, July 2008, <<https://www.rfc-editor.org/info/rfc5250>>.

- [RFC5286] Atlas, A., Ed. and A. Zinin, Ed., "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, DOI 10.17487/RFC5286, September 2008, <<https://www.rfc-editor.org/info/rfc5286>>.
- [RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, DOI 10.17487/RFC5714, January 2010, <<https://www.rfc-editor.org/info/rfc5714>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, DOI 10.17487/RFC7356, September 2014, <<https://www.rfc-editor.org/info/rfc7356>>.
- [RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <<https://www.rfc-editor.org/info/rfc7490>>.
- [RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, DOI 10.17487/RFC7684, November 2015, <<https://www.rfc-editor.org/info/rfc7684>>.
- [RFC7770] Lindem, A., Ed., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 7770, DOI 10.17487/RFC7770, February 2016, <<https://www.rfc-editor.org/info/rfc7770>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [RFC8556] Rosen, E., Ed., Sivakumar, M., Przygienda, T., Aldrin, S., and A. Dolganow, "Multicast VPN Using Bit Index Explicit Replication (BIER)", RFC 8556, DOI 10.17487/RFC8556, April 2019, <<https://www.rfc-editor.org/info/rfc8556>>.

7.2. Informative References

- [I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Decraene, B.,
Francois, P., Voyer, D., Clad, F., and P. Camarillo,
"Topology Independent Fast Reroute using Segment Routing",
draft-ietf-rtgwg-segment-routing-ti-lfa-04 (work in
progress), August 2020.
- [I-D.ietf-spring-segment-protection-sr-te-paths]
Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu,
"Segment Protection for SR-TE Paths", draft-ietf-spring-
segment-protection-sr-te-paths-00 (work in progress),
September 2020.
- [RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A.,
Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation
for Bit Index Explicit Replication (BIER) in MPLS and Non-
MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January
2018, <<https://www.rfc-editor.org/info/rfc8296>>.
- [RFC8401] Ginsberg, L., Ed., Przygienda, T., Aldrin, S., and Z.
Zhang, "Bit Index Explicit Replication (BIER) Support via
IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018,
<<https://www.rfc-editor.org/info/rfc8401>>.
- [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>>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA
USA

Email: Huaimo.chen@futurewei.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing, 102209
China

Email: wangaj3@chinatelecom.cn

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring MD 20904
USA

Phone: 301 502-1347
Email: gyan.s.mishra@verizon.com

Yanhe Fan
Casa Systems
USA

Email: yfan@casa-systems.com

Lei Liu
Fujitsu

USA

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks

McLean, VA
USA

Email: xufeng.liu.ietf@gmail.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 1, 2021

M. McBride
Futurewei
J. Xie
X. Geng
S. Dhanaraj
Huawei
R. Asati
Cisco
Y. Zhu
China Telecom
G. Mishra
Verizon Inc.
Z. Zhang
Juniper
September 28, 2020

BIER IPv6 Requirements
draft-ietf-bier-ipv6-requirements-09

Abstract

There have been several proposed solutions with BIER being used in IPv6. But there hasn't been a document which describes the problem and lists the requirements. The goal of this document is to describe the general BIER IPv6 encapsulation problem and detail solution requirements, thereby assisting the working group in the development of acceptable solutions.

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 April 1, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Requirements Language	3
1.2. Terminology	3
2. Problem Statement	3
3. Requirements	4
3.1. Mandatory Requirements	4
3.1.1. Support various L2 link types	4
3.1.2. Support BIER architecture	4
3.1.3. Support deployment with Non-BFR routers	4
3.1.4. Support OAM	5
3.2. Optional Requirements	5
3.2.1. Support Fragmentation	5
3.2.2. Support IPSEC ESP	5
4. IANA Considerations	5
5. Security Considerations	6
6. Acknowledgement	6
7. Normative References	6
Authors' Addresses	7

1. Introduction

Bit Index Explicit Replication (BIER) [RFC8279] is an architecture that provides optimal multicast forwarding, without requiring intermediate routers to maintain per-flow state, through the use of a multicast-specific BIER header. [RFC8296] defines two types of BIER encapsulation: one is BIER MPLS encapsulation for MPLS environments, the other is non-MPLS BIER encapsulation to run without MPLS. This document describes non-MPLS BIER encapsulation in IPv6 environments. We explain the requirements of transporting multicast flow overlay payload through an IPv6 network underlay using BIER. The solutions

may use IPv6 forwarding plane and may include IPv6 encapsulation and/or generic IPv6 tunnelling.

1.1. 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 RFC 2119 [RFC2119].

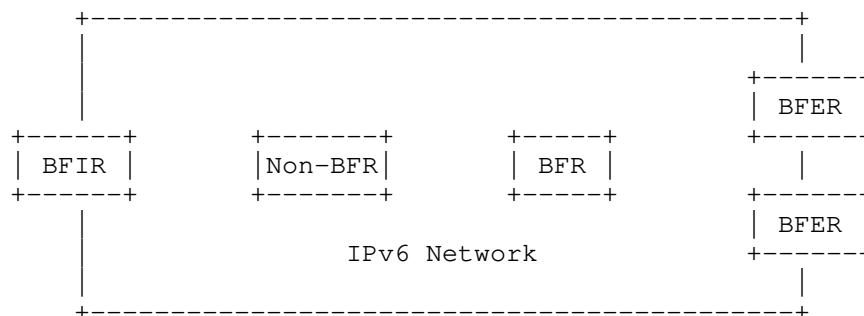
1.2. Terminology

- o BIER: Bit Index Explicit Replication. Provides optimal multicast forwarding through adding a BIER header and removing state in intermediate routers.

2. Problem Statement

The problem is how to transport multicast packets, with non-MPLS BIER encapsulation, in an IPv6 environment. We need to determine where to put the BIER header in this IPv6 environment. With IPv6 encapsulation being increasingly used for unicast services, such as VPN or L2VPN, it may be desirable to have IPv6 encapsulation also used in BIER deployments for multicast services such as MVPN. It may also be desirable to not use IPv6 encapsulation except when IPv6 tunneling (native or GRE/UDP-like) is used to transport BIER packets over BIER-incapable routers.

Below is a simple scenario that needs BIER IPv6-based forwarding:



This scenario depicts the need to replicate BIER packets from a BFIR to BFERs across an IPv6 Service Provider core. Inside the IPv6 network, the BIER header is used to direct the packet from one BFR to the next BFRs, and either a IPv6 header or an L2/tunnel header is used to provide reachability between BFRs. The IPv6 environment may include a variety of link types, may be entirely IPv6, or may be dual stack. There may be cases where not all routers are BFR capable in

the IPv6 environment but still want to deploy BIER. Regardless of the environment, the problem is to deploy BIER, with non-MPLS BIER encapsulation, in an IPv6 network.

3. Requirements

There are several suggested requirements for BIER IPv6 solutions.

In this document, the requirements are divided into two levels: Mandatory and Optional. The requirement levels are determined based on the following factors:

If the requirement is required for a feature that is likely to be a potential deployment, the requirement level will be considered mandatory.

If the impact of not implementing the requirement may block BIER from been deployed, the requirement level will be considered mandatory.

3.1. Mandatory Requirements

Considering that these mandatory requirements are all well-known to the working group, and practical in normal deployment, they will be listed without a detailed description.

3.1.1. Support various L2 link types

The solution should support various kinds of L2 data link types.

3.1.2. Support BIER architecture

The solution must support the BIER architecture.

Supporting different multicast flow overlays, multiple sub-domains, multi-topologies, multiple sets, multiple Bit String Lengths, and deterministic ECMP are considered essential functions of BIER and need to be supported.

3.1.3. Support deployment with Non-BFR routers

The solution must support deployments with BIER-incapable routers. This is beneficial to the deployment of BIER, especially in early deployments when some routers do not support BIER forwarding but support IPv6 forwarding.

3.1.4. Support OAM

BIER OAM tools like [I-D.ietf-bier-ping] and [I-D.ietf-bier-pmmm-oam] should be supported, either directly using existing methods, or by specifying a new method for the same functionality. They are likely to be needed in normal BIER deployment for diagnostics.

3.2. Optional Requirements

The requirements in this section are listed as optional, and each requirement is explained with a detailed scenario. Note that fragmentation and IPSEC ESP are not BIER functions, they are provided by the upper IP layer.

3.2.1. Support Fragmentation

There are some cases where the Fragmentation/Assembly function is needed for BIER to work in an IPv6 network.

For example, a customer IPv6 multicast packet may be 1280 bytes and is required to be transported through an IPv6 network using BIER. Every link of the IPv6 network is no less than the requisite 1280 bytes [RFC8200], but the size of the payload that can be encapsulated in BIER (BIER-MTU) is less than 1280 bytes. In this case, it is not the appropriate action for a BFIR to drop the packet and advertise an MTU to the source [RFC8296]. Instead, some transport mechanism needs to provide the fragmentation and assembly function.

3.2.2. Support IPSEC ESP

There are some cases where the IPSEC ESP function may be needed to transport c-multicast packets through an IPv6 network with confidentiality using BIER technology.

A service provider may want to provide additional security SLA to its customer to ensure that the unencrypted c-multicast packet is not altered in the service provider's network. In this case, if the BIER technology is preferred for the multicast service, BIER with IPSEC ESP support may be a candidate solution. On the other hand, the traffic protection may be better provided via IPSEC or MACSEC at multicast flow overlay over and beyond the BIER domain.

4. IANA Considerations

Some BIER IPv6 encapsulation proposals do not require any action from IANA while other proposals require new IPv6 Option codepoints from IPv6 sub-registries, new "Next header" values, or require new IP

Protocol codes. This document, however, does not require anything from IANA.

5. Security Considerations

There are no security issues introduced by this draft.

6. Acknowledgement

Thanks to Eric Rosen for his listed set of initial requirements on the BIER WG mailing list.

7. Normative References

[I-D.ietf-bier-ping]

Nainar, N., Pignataro, C., Akiya, N., Zheng, L., Chen, M., and G. Mirsky, "BIER Ping and Trace", draft-ietf-bier-ping-07 (work in progress), May 2020.

[I-D.ietf-bier-pmmm-oam]

Mirsky, G., Zheng, L., Chen, M., and G. Fioccola, "Performance Measurement (PM) with Marking Method in Bit Index Explicit Replication (BIER) Layer", draft-ietf-bier-pmmm-oam-08 (work in progress), May 2020.

[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>>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

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

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

Authors' Addresses

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Jingrong Xie
Huawei

Email: xiejingrong@huawei.com

Xuesong Geng
Huawei

Email: gengxuesong@huawei.com

Senthil Dhanaraj
Huawei

Email: senthil.dhanaraj@huawei.com

Rajiv Asati
Cisco

Email: rajiva@cisco.com

Yongqing Zhu
China Telecom

Email: zhuyq8@chinatelecom.cn

Gyan Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

Zhaohui Zhang
Juniper

Email: zzhang@juniper.net

Network Working Group
Internet-Draft
Updates: 8296 (if approved)
Intended status: Standards Track
Expires: August 26, 2021

J. Xie
Huawei Technologies
L. Geng
China Mobile
M. McBride
Futurewei
R. Asati
Cisco
S. Dhanaraj
Huawei
Y. Zhu
China Telecom
Z. Qin
China Unicom
M. Shin
LG Uplus
G. Mishra
Verizon Inc.
X. Geng
Huawei
February 22, 2021

Encapsulation for BIER in Non-MPLS IPv6 Networks
draft-xie-bier-ipv6-encapsulation-10

Abstract

This document proposes a BIER IPv6 (BIERv6) encapsulation for Non-MPLS IPv6 Networks using the IPv6 Destination Option extension header. This document updates RFC 8296.

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] and [RFC8174].

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 August 26, 2021.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Terminology	3
3. BIER IPv6 Encapsulation	4
3.1. BIER Option in IPv6 Destination Options Header	4
3.2. Destination Address in BIERv6 Encapsulation	6
3.3. BIERv6 Packet Format	8
4. BIERv6 Packet Processing	9
5. Security Considerations	11
5.1. Intra Domain Deployment	12
5.2. ICMP Error Processing	13
5.3. Security caused by BIER option	13
5.4. Applicability of IPsec	14
6. IANA Considerations	15
6.1. BIER Option Type	15
6.2. End.BIER Function	15
7. Implementation Status	15
8. Acknowledgements	16
9. Contributors	17
10. References	17
10.1. Normative References	17
10.2. Informative References	18
Appendix A. Relationship to BIER Core Standards	19
Appendix B. Extensions to BIER Control-plane Standards	20
Appendix C. Considerations of Using Unicast Address	20

Authors' Addresses	21
--------------------	----

1. Introduction

Bit Index Explicit Replication (BIER) [RFC8279] is an architecture that provides optimal multicast forwarding without requiring intermediate routers to maintain any per-flow state by using a multicast-specific BIER header.

[RFC8296] defines a common BIER Header format for MPLS and Non-MPLS networks. It has defined two types of encapsulation methods using the common BIER Header, (1) BIER encapsulation in MPLS networks, here-in after referred as MPLS BIER Header in this document and (2) BIER encapsulation in Non-MPLS networks, here-in after referred as Non-MPLS BIER Header in this document. [RFC8296] also assigned Ethertype=0xAB37 for Non-MPLS BIER Header packets to be directly carried over the Ethernet links.

This document proposes a BIER IPv6 encapsulation for Non-MPLS IPv6 Networks, defining a method to carry the standard Non-MPLS BIER header (as defined in [RFC8296]) in the native IPv6 header. A new IPv6 Option type - BIER Option is defined to encode the standard Non-MPLS BIER header and this newly defined BIER Option is carried under the Destination Options header of the native IPv6 Header [RFC8200].

The relationship of this document to BIER core standards is listed in Appendix A.

The relevant extensions to BIER Control-plane Standards are listed in Appendix B.

2. Terminology

Readers of this document are assumed to be familiar with the terminology and concepts of the documents listed as Normative References.

The following new terms are used throughout this document:

- o BIERv6 - Bit indexed explicit replication using IPv6 data plane.
- o BIERv6 Domain - A limited-domain using BIERv6 encapsulation as specified in this document for transporting customer multicast packets from one router to multiple destination routers. It is usually managed by a single administrative entity, e.g., a service-provider. It could be a single AS network or a large-scale network that includes multiple ASes. BIER Domain is also used for the same meaning as BIERv6 domain in this document.

- o BIERv6 Option - An Option type carried in IPv6 Destination Options Header (DO header, DOH) which includes the standard Non-MPLS BIER Header. It is in type-length-value (TLV) format. The value portion of the BIERv6 Option TLV, or the BIERv6 Option Data, is in the format of the standard Non-MPLS BIER header. BIER option is also used for the same meaning as BIERv6 option in this document.
- o BIERv6 Header - An IPv6 Header with BIER Option.
- o BIERv6 Packet - An IPv6 packet with BIERv6 Header. An IP/IPv6/Ethernet multicast packet is encapsulated with an outside BIERv6 header and transformed to a BIERv6 packet on the ingress PE (BFIR). BIERv6 packet is transported by the transit routers (BFRs) through a BIERv6 domain towards egress PEs (BFERs). BIERv6 packet is decapsulated by the BFERs, with the original IP/IPv6/Ethernet multicast packet being obtained and forwarded towards the multicast receivers.

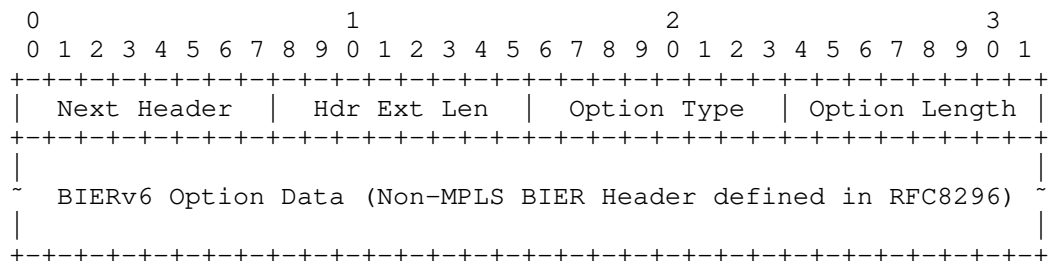
3. BIER IPv6 Encapsulation

3.1. BIER Option in IPv6 Destination Options Header

Destination Options Header and the Options that can be carried under this extension header is defined in [RFC8200]. This document defines a new Option type - BIER Option, to encode the Non-MPLS BIER header. As specified in Section 4.2 [RFC8200], the BIER Option follows type-length-value (TLV) encoding format and the standard Non-MPLS BIER header [RFC8296] is encoded in the value portion of the BIER Option TLV.

This BIER Option MUST be carried only inside the IPv6 Destination Options header and MUST NOT be carried under the Hop-by-Hop Options header.

The BIER Option is encoded in type-length-value (TLV) format as follows:



Next Header 8-bit selector. Identifies the type of header immediately following the Destination Options header.

Hdr Ext Len 8-bit unsigned integer. Length of the Destination Options header in 8-octet units, not including the first 8 octets.

Option Type To be allocated by IANA. See section 6.

Option Length 8-bit unsigned integer. Length of the option, in octets, excluding the Option Type and Option Length fields.

BIERv6 Option Data The BIERv6 Option Data contains the Non-MPLS BIER Header defined in RFC8296. Fields in the Non-MPLS BIER Header MUST be encoded as below.

BIFT-id: The BIFT-id is a domain-wide unique value in Non-MPLS IPv6 encapsulation. See Section 2.2 of RFC 8296.

TC: SHOULD be set to binary value 000 upon transmission and MUST be ignored upon reception. See Section 2.2 of RFC 8296.

S bit: SHOULD be set to 1 upon transmission, and MUST be ignored upon reception. See Section 2.2 of RFC 8296.

TTL: MUST be set to a value larger than 0 upon encapsulation, and SHOULD decrease by 1 by a BFR when forwarding a BIERv6 packet to a BFR adjacency. If the incoming TTL is 0, the packet is considered to be "expired". See Section 2.1.1.2 of RFC 8296.

Nibble: SHOULD be set to 0000 upon transmission, and MUST be ignored upon reception. See Section 2.2 of RFC 8296.

Ver: MUST be set to 0 upon transmission, and MUST be discarded when it is not 0 upon reception. See Section 2.2 of RFC 8296.

BSL: See Section 2.1.2 of RFC 8296.

Entropy: See Section 2.1.2 of RFC 8296.

OAM: See Section 2.1.2 of RFC 8296.

Rsv: See Section 2.1.2 of RFC 8296.

DSCP: SHOULD be set to binary value 000000 upon transmission and MUST be ignored upon reception. In BIERv6 encapsulation, uses Traffic Class field of IPv6 header instead.

Proto: SHOULD be set to 0 upon transmission and be ignored upon reception. In BIERv6 encapsulation, the functionality of this 6-bit Proto field is replaced by the Next Header field in Destination Options header or the last IPv6 extension header to indicate the type of the payload. This updates section 2.1.2 of [RFC8296] about Proto definition. Next Header value in BIERv6 encapsulation for common usage includes:

Value 4 for IPv4 packet as BIERv6 payload.

Value 41 for IPv6 packet as BIERv6 payload.

Value 143 for Ethernet packet as BIERv6 payload.

Multicast VPN (MVPN) service is considered as part of the BIER layering mode defined in [RFC8279], and should be supported by BIERv6 encapsulation. [I-D.xie-bier-ipv6-mvpn] illustrates how MVPN is supported in BIERv6 encapsulation without using this Proto field.

BIER-PING [I-D.ietf-bier-ping] is considered a useful function of the BIER architecture, and should be supported by BIERv6 encapsulation. How BIER-PING is supported in BIERv6 encapsulation without using this Proto field is outside the scope of this document.

BFIR-id: See Section 2.1.2 of RFC 8296.

BitString: See Section 2.1.2 of RFC 8296.

3.2. Destination Address in BIERv6 Encapsulation

When a BIERv6 packet is replicated to a next hop BFR, an unicast address of the next hop BFR is used as the destination address of the BIERv6 packet. Considerations of using unicast (or multicast) address is listed in Appendix C.

The unicast address used in BIERv6 packet targeting a BFR SHOULD be advertised as part of the BIER IPv6 Encapsulation. When a BFR advertises the BIER information with BIERv6 encapsulation capability, an IPv6 unicast address of this BFR MUST be selected specifically for BIERv6 packet forwarding. Locally this "BIER Specific" IPv6 address is initialized in FIB with a flag of "BIER specific handling", represented as End.BIER function.

If a BFR belongs to more than one sub-domain, it may (though it need not) have a different End.BIER in each sub-domain. If different End.BIER is used for each sub-domain, implementation SHOULD support

verifying the DA of a BIERv6 packet is the End.BIER address bound by the sub-domain of the packet.

For security deployment of BIERv6, the End.BIER address(es) is required to be allocated from an IPv6 address block, and the IPv6 address block is used for domain boundary security policy. See section 5.1 of this document for such security policy. Such kind of security policy using IPv6 address block follows the paradigm settled by the [RFC8754] section 5.

Deployment of BIERv6 in SRv6 network is allowed. In this case, the BIERv6 domain is the same as SRv6 domain, and the End.BIER address is allocated from the locator of SRv6.

To better understand the configuration mode of End.BIER address in BIERv6, [I-D.geng-bier-bierv6-yang] could be referenced.

For the convenience of such co-existence of BIERv6 and SRv6, the indication of End.BIER or "BIER specific handling" in FIB shares the same space as SRv6 Endpoints Behaviors defined in [I-D.ietf-spring-srv6-network-programming].

The following is an example pseudo-code of the End.BIER function:

```
1. IF NH = 60 and HopLimit > 0                                ;;Ref1
2.   IF (OptType1 = BIER) and (OptLength1 = HdrExtLen*8 + 4) ;;Ref2
3.     Lookup the BIER Header inside the BIER option TLV.
4.     Forward via the matched entry.
5.   ELSE                                                        ;;Ref3
6.     Drop the packet and end the process.
7. ELSE IF NH=ICMPv6 or (NH=60 and Dest_NH=ICMPv6)             ;;Ref4
8.   Send to CPU.
9. ELSE                                                        ;;Ref5
10.  Drop the packet.
```

Ref1: Destination options header follows the IPv6 header directly and HopLimit is bigger than zero.

Ref2: The first TLV is BIER type and is the only TLV present in Destination options header.

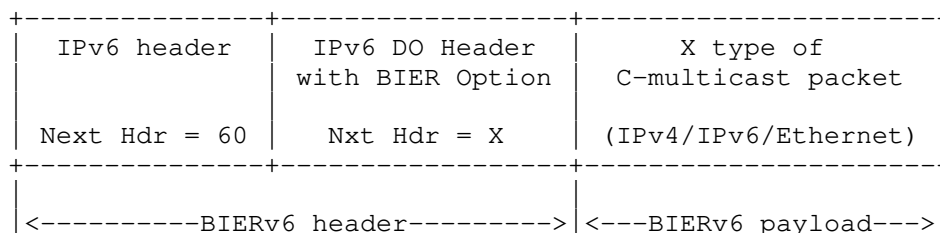
Ref3/Ref5: Undesired packet is dropped because the destination address is the BIER specific IPv6 address (End.BIER function).

Ref4: An ICMPv6 packet using End.BIER as destination address.

3.3. BIERv6 Packet Format

As a multicast packet enters the BIER domain in a Non-MPLS IPv6 network, the multicast packet will be encapsulated with BIERv6 Header by the Ingress BFR (BFIR).

Typically a BIERv6 header would contain the Destination Options Header as the only Extensions Header besides IPv6 Header, as depicted in the below figure.



Format of the multicast packet with BIERv6 encapsulation carrying other extension headers along with Destination Options extension header is required to follow general recommendations of [RFC8200] and examples in other RFCs. [RFC6275] introduces how the order should be when other extension headers carries along with Home address option in a destination options header. Similar to this example, this document requires the Destination Options Header carrying the BIER option MUST be placed as follows:

- o After the routing header, if that header is present
- o Before the Fragment Header, if that header is present
- o Before the AH Header or ESP Header, if either one of those headers is present

Source Address field in the IPv6 header MUST be a routable IPv6 unicast address of the BFIR in any case.

BFIR encodes the BIERv6 header in the above mentioned encapsulation format and forwards the BIERv6 packet to the nexthop BFR following the local BIFT table.

BFRs in the IPv6 network, processes and replicates the packets towards the BFRs using the local BIFT table. The BitString field in the BIERv6 Option Data may be changed by the BFRs as they replicate the packet. BFRs MUST follow the procedures defined in section 3.1 as they modify the other fields in the BIERv6 Option Data. The source address in the IPv6 header MUST NOT be modified by the BFRs.

4. BIERv6 Packet Processing

When a multicast packet enters the BIER domain, the Ingress BFR (BFIR) encapsulates the multicast packet with a BIERv6 Header, transforming it to a BIERv6 packet. The BIERv6 header includes an IPv6 header and a BIERv6 Option in IPv6 Destination Options Header. Source Address field in the IPv6 header MUST be set to a routable IPv6 unicast address of the BFIR. Destination Address field in the IPv6 header is set to the End.BIER address of the next-hop BFR the BIERv6 packet replicating to, no matter next-hop BFR is directly connected (one-hop) or not directly connected (multi-hop).

Upon receiving an BIERv6 packet, the BFR processes the IPv6 header first. This is the general procedure of IPv6.

If the IPv6 Destination address is an End.BIER IPv6 unicast address of this BFR, a 'BIER Specific Handling' indication will be obtained by the preceding Unicast DA lookup (FIB lookup). The BIER option, if exists, will be checked to decide which neighbor(s) to replicate the BIERv6 packet to.

It is a local behavior to handle the combination of extension headers, options and the BIER option(s) in destination options header when a 'BIER Specific Handling' indication is got by the preceding FIB lookup. Early deployment of BIERv6 may require there is only one BIER option TLV in the destination options header followed the IPv6 header. How other extension headers or more BIER option TLVs in a BIERv6 packet is handled is outside the scope of this document.

A packet having a 'BIER Specific Handling' indication but not having a BIER option is supposed to be a wrong packet or an ICMPv6 packet, and the process can be referred to the example in section 3.2.

A packet not having a 'BIER Specific Handling' indication but having a BIER option SHOULD be processed normally as unicast forwarding procedures, which may be a behavior of drop, or send to CPU, or other behaviors in existing implementations.

The Destination Address field in the IPv6 Header MUST change to the nexthop BFR's End.BIER Unicast address in BIERv6.

The Hop Limit field of IPv6 header MUST decrease by 1 when sending packets to a BFR neighbor, while the TTL in the BIER header MUST be unchanged on a Non-BIER router, or decrease by 1 on a BFR.

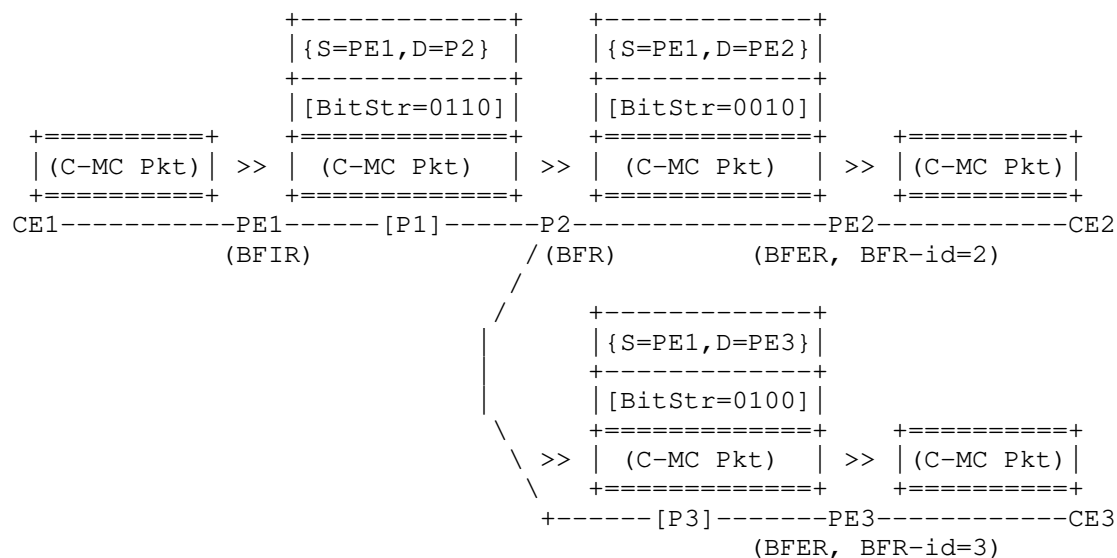
The BitString in the BIER header in the Destination Options Header may change when sending packets to a neighbor. Such change of BitString MUST be aligned with the procedure defined in RFC8279.

Because of the requirement to change the content of the option when forwarding BIERv6 packet, the BIER option type should have chg flag 1 per section 4.2 of RFC8200.

The procedures applies normally if a bit corresponding to the self bfr-id is set in the BitString field of the BIERv6 Option Data of the BIERv6 packet. The node is considered to be an Egress BFR (BFER) in this case. The BFER removes the BIERv6 header, including the IPv6 header and the Destination Options header, and copies the packet to the multicast flow overlay. The egress VRF of a packet may be determined by a further lookup on the IPv6 source address instead of the upstream-assigned MPLS Label as described in [RFC8556].

The Fragment Header, AH Header or ESP Header, if exists after the BIER options header, can be processed on BFER only as part of the multicast flow overlay process.

The following diagram shows the whole progression of the multicast packet as it enters the BIERv6 domain on PE1, and leaves the BIERv6 domain on PE2 and PE3.



{S=PE1,D=PE2}: Source address and Destination address in IPv6 header.

[BitStr=0110]: BitString value in IPv6 DO Header.

(C-MC Pkt): Customer MultiCast packet.

- o PE1 is Provider Edge router, acting as BFIR.

- o P2 is Provider Core router, acting as BFR.
- o P1 and P3 are IPv6 routers, acting as Non-BFR.
- o PE2 and PE3 are Provider Edge routers, acting as BFER.
- o CE1 and CE2 are Customer Edge routers.

5. Security Considerations

BIER IPv6 encapsulation provides a new encapsulation based on IPv6 and BIER to transport multicast data packet in a BIER domain. The BIER domain can be a single IGP area, an anonymous system (AS) with multiple IGP areas, or multiple anonymous systems (ASes) operated by a network operator. A single BIER Sub-domain may be deployed through the whole BIER Domain, as illustrated in [I-D.geng-bier-ipv6-inter-domain].

This section reviews security considerations related to the BIER IPv6 encapsulation, based on security considerations of [RFC8279], [RFC8296], and other documents related to IPv6 extension.

It is expected that all nodes in a BIER IPv6 domain are managed by the same administrative entity. BIER-encapsulated packets should generally not be accepted from untrusted interfaces or tunnels. For example, an operator may wish to have a policy of accepting BIER-encapsulated packets only from interfaces to trusted routers, and not from customer-facing interfaces. See section 5.1 for normal Intra domain deployment.

For applications that require a BFR to accept a BIER-encapsulated packet from an interface to a system that is not controlled by the network operator, the security considerations of [RFC8296] apply.

BIER IPv6 encapsulation may cause ICMP packet sent to BFIR and cause security problems. See section 5.2 for ICMP related problems.

This document introduces a new option used in IPv6 Destination Options Header, together with the special-use IPv6 address called End.BIER in IPv6 destination address for BIER IPv6 forwarding. However the option newly introduced may be wrongly used with normal IPv6 destination address. See section 5.3 for problems introduced by the new IPv6 option with normal IPv6 destination address.

If the multicast data packet of a BIERv6 packet is altered by an intermediate router, contents of the multicast data packet will be damaged. BIER IPv6 encapsulation provides the ability of IPsec to

ensure the confidentiality or integrity for multicast data packet. See section 5.4 for this security problem.

If the BIERv6 encapsulation of a particular packet specifies a BitString (together with SI) other than the one intended by the BFIR, the packet is likely to be misdelivered. Some modifications of the BIER encapsulation, e.g., setting every bit in the BitString, may result in denial-of-service (DoS) attacks. This kind of DoS attack is a challenge not only in BIERv6 but also in BIER as specified in [RFC8279] and [RFC8296], as the BitString is required to change on BFR per the BIER forwarding procedures. This document does not provide new mechanisms to improve this kind of weakness.

A BIER router accepts and uses the End.BIER IPv6 address to construct BIFT only when the IPv6 address is configured explicitly, or received from a router via control-plane protocols. The received information is validated using existing authentication and security mechanisms of the control-plane protocols. BIER IPv6 encapsulation does not define any additional security mechanism in existing control-plane protocols, and it inherits any security considerations that apply to the control-plane protocols.

5.1. Intra Domain Deployment

Generally nodes outside the BIER Domain are not trusted: they cannot directly use the End.BIER of the domain. This is enforced by two levels of access control lists:

1. Any packet entering the BIER Domain and destined to an End.BIER IPv6 Address within the BIER Domain is dropped. This may be realized with the following logic. Other methods with equivalent outcome are considered compliant:

- * allocate all the End.BIER IPv6 Address from a block S/s

- * configure each external interface of each edge node of the domain with an inbound infrastructure access list (IACL) which drops any incoming packet with a destination address in S/s

- * Failure to implement this method of ingress filtering exposes the BIER Domain to BIER attacks as described and referenced in [RFC8296].

2. The distributed protection in #1 is complemented with per node protection, dropping packets to End.BIER IPv6 Address from source addresses outside the BIER Domain. This may be realized with the following logic. Other methods with equivalent outcome are considered compliant:

- * assign all interface addresses from prefix A/a
- * assign all the IPv6 addresses used as source address of BIER IPv6 packets from a block B/b
- * at node k, all End.BIER IPv6 addresses local to k are assigned from prefix Sk/sk
- * configure each internal interface of each BIER node k in the BIER Domain with an inbound IACL which drops any incoming packet with a destination address in Sk/sk if the source address is not in A/a or B/b.

For simplicity of deployment, a configuration of IACL effective for all interfaces can be provided by a router. Such IACL can be referred to as global IACL(GIACL). Each BIER node k then simply config a GIACL which drops any incoming packet with a destination address in Sk/sk if the source address is not in A/a or B/b for the intra-domain deployment mode.

5.2. ICMP Error Processing

The BIERv6 BFR does not send ICMP error messages to the source address of a BIERv6 packet, there is still chance that Non-BFR routers send ICMP error messages to source nodes within the BIER Domain.

A large number of ICMP may be elicited and sent to a BFIR router, in case when a BIERv6 packet is filled with wrong Hop Limit, either error or malfeasance. A rate-limiting of ICMP packet should be implemented on each BFR.

The ingress node can take note of the fact that it is getting, in response to BIER IPv6 packet, one or more ICMP error packets. By default, the reception of such a packets MUST be countered and logged. However, it is possible for such log entries to be "false positives" that generate a lot of "noise" in the log; therefore, implementations SHOULD have a knob to disable this logging.

5.3. Security caused by BIER option

This document introduces a new option used in IPv6 Destination Options Header. An IPv6 packet with a normal IPv6 address of a router (e.g. loopback IPv6 address of the router) as destination address will possibly carry a BIER option.

For a router incapable of BIERv6, such BIERv6 packet will not be processed by the procedure described in this document, but be

processed as normal IPv6 packet with unknown option, and the existing security considerations for handling IPv6 options apply. Possible way of handling IPv6 packets with BIER option may be send to CPU for slow path processing, with rate-limiting, or be discarded according to the local policy.

For a router capable of BIERv6, such BIERv6 packet MUST NOT be forwarded, but should be processed as a normal IPv6 packet with unknown option, or additionally and optionally be countered and logged if the router is capable of doing so.

5.4. Applicability of IPsec

IPsec [RFC4301] uses two protocols to provide traffic security services -- Authentication Header (AH) [RFC4302] and Encapsulating Security Payload (ESP) [RFC4303]. Each protocol supports two modes of use: transport mode and tunnel mode. IPsec support both unicast and multicast. IPsec implementations MUST support ESP and MAY support AH.

This document assume IPsec working in tunnel mode with inner IPv4 or IPv6 multicast packet encapsulated in outer BIERv6 header and IPsec header(s).

IPsec used with BIER IPv6 encapsulation to ensure that a BIER payload is not altered while in transit between BFIR and BFERs. If a BFR in between BFIR and BFERs is compromised, there is no way to prevent the compromised BFR from making illegitimate modifications to the BIER payload or to prevent it from misforwarding or misdelivering the BIER-encapsulated packet, but the BFERs will detect the illegitimate modifications to the BIER Payload (or the inner multicast data packet). This could provide cryptographic integrity protection for multicast data transport. This capability of IPsec comes from the design that, the destination options header carrying the BIER header is located before the AH or ESP and the BFR routers in between BFIR and BFERs can process the BIER header without aware of AH or ESP.

For ESP, the Integrity Check Value (ICV) is computed over the ESP header, Payload, and ESP trailer fields. It doesn't require the IP or extension header for ICV calculating, and thus the change of DA and BIER option data does not affect the function of ESP.

For AH, the Integrity Check Value (ICV) is computed over the IP or extension header fields before the AH header, the AH header, and the Payload. The IPv6 DA is immutable for unicast traffic in AH, and the change of DA in BIER IPv6 forwarding for multicast traffic is incompatible to this rule. How AH is extended to support multicast

traffic transporting through BIER IPv6 encapsulation is outside the scope of this document.

The detailed control-plane for BIER IPv6 encapsulation IPsec function is outside the scope of the document. Internet Key Exchange Protocol Version 2 (IKEv2) [RFC7296] and Group Security Association (GSA) [RFC5374] can be referred to for further studying.

6. IANA Considerations

6.1. BIER Option Type

Allocation is expected from IANA for a BIER Option Type codepoint from the "Destination Options and Hop-by-Hop Options" sub-registry of the "Internet Protocol Version 6 (IPv6) Parameters" registry.

Hex Value	act	chg	rest	Description	Reference
TBD	01	1	TBD	BIER Option	This draft

6.2. End.BIER Function

Allocation is expected from IANA for an End.BIER function codepoint from the "SRv6 Endpoint Behaviors" sub-registry. The value 60 is suggested.

Value	Hex	Endpoint function	Reference
TBD	TBD	End.BIER	This draft

7. Implementation Status

Implementation of BIERv6 as described in this document and related drafts defined in appendix B has been operational.

In February 2021, China Unicom successfully validated this BIERv6 solution over its Beijing Metro network.

The BIERv6 test network contains the following network devices:

STB, ONT, OLT, BFER,non-BFR, BFR, BFIR, Switch, CR (Core Router, video flow input node)

where BIERv6 capable nodes include:

- o BFER: Huawei ATN980C
- o BFR&BFIR: Huawei NE40E

and IPv6 capable nodes include:

- o non-BFR: Cisco ASR 9006

BIERv6 test is composed of the following scenarios:

- o Intra-AS BIERv6: mechanisms defined in this document and [I-D.xie-bier-ipv6-isis-extension] are validated.
- o Inter-AS BIERv6: mechanisms defined in [I-D.geng-bier-ipv6-inter-domain] are validated.
- o IPv4 multicast traffic over BIERv6 MVPN. Mechanisms defined in [I-D.xie-bier-ipv6-mvpn] are validated.
- o Deployment of Intra-AS and Inter-AS BIERv6 with non-BIERv6 capable intermediate nodes, where these nodes only need to be IPv6 capable.
- o BIERv6 Ping
- o Deterministic convergence as indicated in [RFC8279], where intermediate link or BFR node fails.
- o Service reliability where BIERv6 source side link or BFIR node fails.
- o Service reliability where BIERv6 receiver side link or BFER node fails.

8. Acknowledgements

The authors would like to thank Stig Venaas for his valuable comments. Thanks IJsbrand Wijnands, Greg Shepherd, Tony Przygienda, Toerless Eckert, Jeffrey Zhang, Pascal Thubert for the helpful comments to improve this document.

Thanks Aijun Wang for comments about BIER OAM function in BIER IPv6 encapsulation.

Thanks Mach Chen for review and suggestions about BIER-PING function in BIER IPv6 encapsulation.

9. Contributors

Gang Yan

Huawei Technologies

China

Email: yangang@huawei.com

Yang(Yolanda) Xia

Huawei Technologies

China

Email: yolanda.xia@huawei.com

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <<https://www.rfc-editor.org/info/rfc4301>>.
- [RFC4302] Kent, S., "IP Authentication Header", RFC 4302, DOI 10.17487/RFC4302, December 2005, <<https://www.rfc-editor.org/info/rfc4302>>.
- [RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, DOI 10.17487/RFC4303, December 2005, <<https://www.rfc-editor.org/info/rfc4303>>.
- [RFC5374] Weis, B., Gross, G., and D. Ignjatic, "Multicast Extensions to the Security Architecture for the Internet Protocol", RFC 5374, DOI 10.17487/RFC5374, November 2008, <<https://www.rfc-editor.org/info/rfc5374>>.
- [RFC6275] Perkins, C., Ed., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, DOI 10.17487/RFC6275, July 2011, <<https://www.rfc-editor.org/info/rfc6275>>.

- [RFC7296] Kaufman, C., Hoffman, P., Nir, Y., Eronen, P., and T. Kivinen, "Internet Key Exchange Protocol Version 2 (IKEv2)", STD 79, RFC 7296, DOI 10.17487/RFC7296, October 2014, <<https://www.rfc-editor.org/info/rfc7296>>.
- [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>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January 2018, <<https://www.rfc-editor.org/info/rfc8296>>.
- [RFC8401] Ginsberg, L., Ed., Przygienda, T., Aldrin, S., and Z. Zhang, "Bit Index Explicit Replication (BIER) Support via IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018, <<https://www.rfc-editor.org/info/rfc8401>>.
- [RFC8556] Rosen, E., Ed., Sivakumar, M., Przygienda, T., Aldrin, S., and A. Dolganow, "Multicast VPN Using Bit Index Explicit Replication (BIER)", RFC 8556, DOI 10.17487/RFC8556, April 2019, <<https://www.rfc-editor.org/info/rfc8556>>.
- [RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.

10.2. Informative References

- [I-D.geng-bier-bierv6-yang]
Geng, X., Qin, Z., and F. Zheng, "YANG Data Model for Bierv6", draft-geng-bier-bierv6-yang-00 (work in progress), June 2020.

[I-D.geng-bier-ipv6-inter-domain]

Geng, L., Xie, J., McBride, M., Yan, G., and X. Geng, "Inter-Domain Multicast Deployment using BIERv6", draft-geng-bier-ipv6-inter-domain-02 (work in progress), October 2020.

[I-D.ietf-bier-ipv6-requirements]

McBride, M., Xie, J., Geng, X., Dhanaraj, S., Asati, R., Zhu, Y., Mishra, G., and Z. Zhang, "BIER IPv6 Requirements", draft-ietf-bier-ipv6-requirements-09 (work in progress), September 2020.

[I-D.ietf-bier-ping]

Nainar, N., Pignataro, C., Akiya, N., Zheng, L., Chen, M., and G. Mirsky, "BIER Ping and Trace", draft-ietf-bier-ping-07 (work in progress), May 2020.

[I-D.ietf-spring-srv6-network-programming]

Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-ietf-spring-srv6-network-programming-28 (work in progress), December 2020.

[I-D.xie-bier-ipv6-isis-extension]

Xie, J., Wang, A., Yan, G., Dhanaraj, S., and X. Geng, "BIER IPv6 Encapsulation (BIERv6) Support via IS-IS", draft-xie-bier-ipv6-isis-extension-02 (work in progress), October 2020.

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

Xie, J., McBride, M., Dhanaraj, S., Geng, L., and G. Mishra, "Use of BIER IPv6 Encapsulation (BIERv6) for Multicast VPN in IPv6 networks", draft-xie-bier-ipv6-mvpn-03 (work in progress), October 2020.

Appendix A. Relationship to BIER Core Standards

The BIER architecture [RFC8279] is inherited in this BIERv6 proposal, and the layering mode of BIER architecture is fully supported with some necessary extension to the data plane as well as the control plane standards.

The focus of this document is BIERv6 data plane, including the BIERv6 encapsulation and packet forwarding procedures. The common BIER header encoding [RFC8296] is maximum reused in this BIERv6 proposal.

To better understand the overall BIER IPv6 problem space and requirements, refer to [I-D.ietf-bier-ipv6-requirements].

Appendix B. Extensions to BIER Control-plane Standards

The relevant control-plane documents that have done or still to be done are listed below.

- o Based on [RFC8401], IS-IS extension is defined in [I-D.xie-bier-ipv6-isis-extension] for intra-AS BIERv6 information advertisement and BIRT/BIFT building.
- o OSPFv3 extension for intra-AS BIERv6 information advertisement and BIRT/BIFT building is to be defined.
- o Based on this BIERv6 encapsulation, [I-D.geng-bier-ipv6-inter-domain] illustrates how inter-AS BIRT/BIFT are built and how inter-AS multicast deployment is supported.
- o BGP extension for inter-AS BIERv6 information advertisement and BIRT/BIFT building is to be defined.
- o Based on [RFC8556], BGP-MVPN using BIERv6 encapsulation is defined in [I-D.xie-bier-ipv6-mvpn] for multicast service deployment.

Appendix C. Considerations of Using Unicast Address

BIER is generally a hop-by-hop and one-to-many architecture, and thus the IPv6 Destination Address (DA) being a Multicast Address is a way one may think of as an approach for both the two paradigms in BIERv6 encapsulation.

However using a unicast address has the following benefits:

1. Replicating a BIERv6 packet over a non-BIER capable router.
2. Fast rerouting a BIERv6 packet using a unicast by-pass tunnel.
3. Forwarding a BIERv6 packet to one of the many BFR neighbors connected on a LAN without imposing new requirements of snooping on switches.
4. Replicating a BIERv6 packet through an anonymous system(AS) to BFRs in other ASes, as illustrated in [I-D.geng-bier-ipv6-inter-domain].

Some of the above scenarios are assumed part of BIER architecture as described in [RFC8279], and some of them are the scalability aspects for inter-AS stateless multicast this document intends to support. This document intends to fulfil all these requirements (categorized

as multi-hop replication), and proposes to use unicast address for both one-hop replication and multi-hop replication.

Authors' Addresses

Jingrong Xie
Huawei Technologies

Email: xiejingrong@huawei.com

Liang Geng
China Mobile
Beijing 10053

Email: gengliang@chinamobile.com

Mike McBride
Futurewei

Email: mmcbride7@gmail.com

Rajiv Asati
Cisco

Email: rajiva@cisco.com

Senthil Dhanaraj
Huawei

Email: senthil.dhanaraj@huawei.com

Yongqing Zhu
China Telecom

Email: zhuyq8@chinatelecom.cn

Zhuangzhuang Qin
China Unicom

Email: qinzhuangzhuang@chinaunicom.cn

MooChang Shin
LG Uplus

Email: himzzang@lguplus.co.kr

Gyan Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

Xuesong Geng
Huawei

Email: gengxuesong@huawei.com

BIER
Internet-Draft
Intended status: Standards Track
Expires: August 26, 2021

Z. Zhang
ZTE Corporation
Z. Zhang, Ed.
Juniper Networks
I. Wijnands
Individual
M. Mishra
Cisco Systems
H. Bidgoli
Nokia
G. Mishra, Ed.
Verizon
February 22, 2021

Supporting BIER in IPv6 Networks (BIERin6)
draft-zhang-bier-bierin6-09

Abstract

BIER is a new architecture for the forwarding of multicast data packets without requiring per-flow state inside the network. This document describes how the existing BIER encapsulation specified in RFC 8296 works in an IPv6 non-MPLS network, referred to as BIERin6. Specifically, like in an IPv4 network, BIER can work over L2 links directly or over tunnels. In case of IPv6 tunneling, a new IP "Next Header" type is to be assigned for BIER.

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 August 26, 2021.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. BIER over L2/Tunnels	3
1.2. Considerations of Requirements for BIER in IPv6 Networks	3
2. IPv6 Header	5
2.1. IPv6 Options Considerations	5
3. BIER Header	6
4. IPv6 Encapsulation Advertisement	6
4.1. Format	6
4.2. Inter-area prefix redistribution	7
5. IANA Considerations	7
6. Security Considerations	7
7. Acknowledgement	7
8. References	7
8.1. Normative References	8
8.2. Informative References	8
Authors' Addresses	10

1. Introduction

BIER [RFC8279] is a new architecture for the forwarding of multicast data packets. It provides optimal forwarding through a "multicast domain" and it does not precondition construction of a multicast distribution tree, nor does it require intermediate nodes to maintain any per-flow state.

This document specifies non-MPLS BIER forwarding in an IPv6 [RFC8200] environment, referred to as BIERin6, using non-MPLS BIER encapsulation specified in [RFC8296].

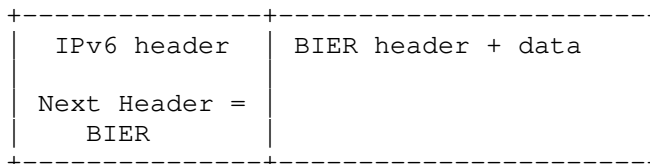
MPLS BIER forwarding in IPv6 is outside the scope of this document.

This document uses terminology defined in [RFC8279] and [RFC8296].

1.1. BIER over L2/Tunnels

[RFC8296] defines the BIER encapsulation format in MPLS and non-MPLS environment. In case of non-MPLS environment, a BIER packet is the payload of an "outer" encapsulation, which has a "next header" codepoint that is set to a value that means "non-MPLS BIER". This "BIER over L2/Tunnel" model can be used as is in an IPv6 non-mpls environment, and is referred to as BIERin6.

If a BFR needs to tunnel BIER packets to another BFR, e.g. per [RFC8279] Section 6.9, while any type of tunnel will work, for best efficiency native IPv6 encapsulation can be used with the destination address being the downstream BFR and the Next Header field set to a to-be-assigned value for "non-MPLS BIER".



Between two directly connected BFRs, a BIER header can directly follow link layer header, e.g., an Ethernet header (with the Ethertype set to 0xAB37). Optionally, IPv6 encapsulation can be used even between directly connected BFRs (i.e. one-hop IPv6 tunneling) in the following two cases:

- o An operator mandates all traffic to be carried in IPv6.
- o A BFR does not have BIER support in its "fast forwarding path" and relies on "slow/software forwarding path", e.g. in environments like [RFC7368] where high throughput multicast forwarding performance is not critical.

1.2. Considerations of Requirements for BIER in IPv6 Networks

[draft-ietf-bier-ipv6-requirements] lists mandatory and optional requirements for BIER in IPv6 Networks. As a solution based on the

BIER over L2/tunnel model [RFC8296], BIERin6 satisfies all the mandatory requirements.

For the two optional requirements for fragmentation and Encapsulating Security Payload (ESP), they can be satisfied by one of two ways:

- o IPv6 based fragmentation/ESP: a BFIR encapsulates the payload in IPv6 with fragmentation and/or ESP header, and then the IPv6 packets are treated as BIER payload.
- o Generic Fragmentation/ESP [I-D.zzhang-tsvwg-generic-transport-functions]: a BFIR does generic fragmentation and/or ESP (without using IPv6 encapsulation) and the resulting packets are treated as BIER payload.

Either way, the fragmentation/ESP is handled by a layer outside of BIER and then the resulting packets are treated as BIER payload.

BIERin6 does support SRv6 based overlay services (e.g. MVPN/EVPN). One of the following methods can be used (relevant overlay signaling will be specified separately):

- o An ingress PE (which is a BFIR) can encapsulate customer packets with an IPv6 header (with optional fragmentation and ESP extension headers). The destination address is a multicast locator plus the Fucn/Arg portion that identifies the service. That IPv6 packet is then treated as BIER payload. An egress PE (which is a BFER) uses the standard SRv6 procedures to forward the IPv6 packet that is exposed after the BIER header is decapsulated.
- o Alternatively, since only the destination IPv6 address in the above-mentioned IPv6 header is used for service delimiting purpose, a new value can be assigned for the Proto field in the BIER header to indicate that an IPv6 address (instead of an entire IPv6 header) is added between the BIER header and original payload.

BIERin6 being a solution based on [RFC8279] [RFC8296], ECMP is inherently supported by BFRs using the the 20-bit entropy field in the BIER header for the load balancing hash. When a BIER packet is transported over an IPv6 tunnel, the entropy value is copied into the 20-bit IPv6 Flow Label (instead of using local 5-tuple input key to a hash function to locally generate the stateless 20-bit flow label) so that routers along the tunnel can do ECMP based on Flow Labels. For a router along the tunnel doing deep packet inspection for ECMP purpose, if it understands BIER header it can go past the BIER header to look for the 5-tuple input key to a hash function, otherwise it

stops at the BIER header. In either case the router will not mistake the BIER header as an IP header so no misordering should happen.

BIER has its own OAM functions independent of those related to the underlying links or tunnels. With BIERin6 following the "BIER over L2/tunnel" model, IPv6 OAM function and BIER OAM functions are used independently for their own purposes.

Specifically, BIERin6 works with all of the following OAM methods, or any future methods that are based on the "BIER over L2/tunnel" model:

- o BIER OAM specified in [I-D.ietf-bier-ping]
- o BIER BFD specified in [I-D.ietf-bier-bfd]
- o BIER Performance Measurement specified in [I-D.ietf-bier-pmm-m-oam]
- o BIER Path Maximum Transmission Unit Discovery specified in [I-D.ietf-bier-path-mtu-discovery]
- o BIER IOAM specified in [I-D.xzlnp-bier-ioam]

2. IPv6 Header

Whenever IPv6 encapsulation is used for BIER forwarding, The Next Header field in the IPv6 Header (if there are no extension headers), or the Next Header field in the last extension header is set to TBD, indicating that the payload is a BIER packet.

If the neighbor is directly connected, The destination address in IPv6 header SHOULD be the neighbor's link-local address on this router's outgoing interface, the source destination address SHOULD be this router's link-local address on the outgoing interface, and the IPv6 TTL MUST be set to 1. Otherwise, the destination address SHOULD be the BIER prefix of the BFR neighbor, the source address SHOULD be this router's BIER prefix, and the TTL MUST be large enough to get the packet to the BFR neighbor.

The "Flow label" field in the IPv6 packet SHOULD be copied from the entropy field in the BIER encapsulation.

2.1. IPv6 Options Considerations

For directly connected BIER routers, IPv6 Hop-by-Hop or Destination options are irrelevant and SHOULD NOT be inserted by BIER on the BIERin6 packet. In this case IPv6 header, Next Header field should be set to TBD. Any IPv6 packet arriving on BFRs and BFERs, with multiple extension header where the last extension header has a Next

Header field set to TBD, SHOULD be discard and the node should transmit an ICMP Parameter Problem message to the source of the packet (BFIR) with an ICMP code value of TBD10 ('invalid options for BIERin6').

This also indicates that for disjoint BIER routers using IPv6 encapsulation, there SHOULD NOT be any IPv6 Hop-by-Hop or Destination options be present in a BIERin6 packet. In this case, if additional traffic engineering is required, IPv6 tunneling (i.e. BIERin6 over SRv6) can be implemented.

3. BIER Header

The BIER header MUST be encoded per Section 2.2 of [RFC8296].

The BIFT-id is either encoded per [I-D.ietf-bier-non-mpls-bift-encoding] or per advertised by BFRs, as specified in [I-D.ietf-bier-lsr-ethernet-extensions].

4. IPv6 Encapsulation Advertisement

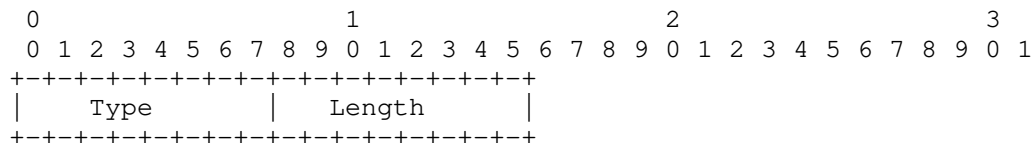
When IPv6 encapsulation is not required between directly connected BFRs, no signaling in addition to that specified in [I-D.ietf-bier-lsr-ethernet-extensions] is needed.

Otherwise, a node that requires IPv6 encapsulation MUST advertise the BIER IPv6 transportation sub-sub-sub-TLV/sub-sub-TLV according to local configuration or policy in the BIER domain to request other BFRs to always use IPv6 encapsulation.

In presence of multiple encapsulation possibilities hop-by-hop it is a matter of local policy which encapsulation is imposed and the receiving router MUST accept all encapsulations that it advertised.

4.1. Format

The BIER IPv6 transportation is a new sub-sub-TLV of BIER Ethernet Encapsulation sub-TLV defined in OSPFv3, and a new sub-sub-sub-TLV of BIER Ethernet Encapsulation sub-sub-TLV defined in ISIS, as per [I-D.ietf-bier-lsr-ethernet-extensions].



- o Type: For OSPF, value TBD1 (prefer 1) is used to indicate it is the IPv6 transportation sub-TLV. For ISIS, value TBD2 (prefer 1) is used to indicate it is the IPv6 transportation sub-sub-TLV.
- o Length: 0.

4.2. Inter-area prefix redistribution

When BFR-prefixes are advertised across IGP areas per [I-D.ietf-bier-lsr-ethernet-extensions] or redistributed across protocol boundaries per [I-D.ietf-bier-prefix-redistribute], the BIER IPv6 transportation sub-sub-TLV or sub-sub-sub-TLV MAY be re-advertised/re-distributed as well.

5. IANA Considerations

IANA is requested to assign a new "BIER" type for "Next Header" in the "Assigned Internet Protocol Numbers" registry.

IANA is requested to assign a new "BIERin6" type for "invalid options" in the "ICMP code value" registry.

IANA is requested to assign a new "IPv6 address" type in the "BIER Next Protocol Identifiers" registry.

IANA is requested to assign a new "BIER IPv6 transportation Sub-sub-TLV" type in the "OSPFv3 BIER Ethernet Encapsulation sub-TLV" Registry.

IANA is requested to set up a new "BIER IPv6 transportation Sub-sub-sub-TLV" type in the "IS-IS BIER Ethernet Encapsulation sub-sub-TLV" Registry.

6. Security Considerations

General IPv6 and BIER security considerations apply.

7. Acknowledgement

The authors would like to thank Tony Przygienda, Nagendra Kumar for their review and valuable comments.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, DOI 10.17487/RFC6437, November 2011, <<https://www.rfc-editor.org/info/rfc6437>>.
- [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>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January 2018, <<https://www.rfc-editor.org/info/rfc8296>>.
- [RFC8401] Ginsberg, L., Ed., Przygienda, T., Aldrin, S., and Z. Zhang, "Bit Index Explicit Replication (BIER) Support via IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018, <<https://www.rfc-editor.org/info/rfc8401>>.

8.2. Informative References

- [I-D.ietf-bier-bar-ipa] Zhang, Z., Przygienda, T., Dolganow, A., Bidgoli, H., Wijnands, I., and A. Gulko, "BIER Underlay Path Calculation Algorithm and Constraints", draft-ietf-bier-bar-ipa-07 (work in progress), September 2020.

[I-D.ietf-bier-bfd]

Xiong, Q., Mirsky, G., hu, f., and C. Liu, "BIER BFD", draft-ietf-bier-bfd-00 (work in progress), November 2020.

[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-07 (work in progress), September 2019.

[I-D.ietf-bier-ipv6-requirements]

McBride, M., Xie, J., Geng, X., Dhanaraj, S., Asati, R., Zhu, Y., Mishra, G., and Z. Zhang, "BIER IPv6 Requirements", draft-ietf-bier-ipv6-requirements-09 (work in progress), September 2020.

[I-D.ietf-bier-lsr-ethernet-extensions]

Dhanaraj, S., Yan, G., Wijnands, I., Psenak, P., Zhang, Z., and J. Xie, "LSR Extensions for BIER over Ethernet", draft-ietf-bier-lsr-ethernet-extensions-02 (work in progress), December 2020.

[I-D.ietf-bier-non-mpls-bift-encoding]

Wijnands, I., Mishra, M., Xu, X., and H. Bidgoli, "An Optional Encoding of the BIFT-id Field in the non-MPLS BIER Encapsulation", draft-ietf-bier-non-mpls-bift-encoding-03 (work in progress), November 2020.

[I-D.ietf-bier-ospfv3-extensions]

Psenak, P., Nainar, N., and I. Wijnands, "OSPFv3 Extensions for BIER", draft-ietf-bier-ospfv3-extensions-03 (work in progress), November 2020.

[I-D.ietf-bier-path-mtu-discovery]

Mirsky, G., Przygienda, T., and A. Dolganow, "Path Maximum Transmission Unit Discovery (PMTUD) for Bit Index Explicit Replication (BIER) Layer", draft-ietf-bier-path-mtu-discovery-09 (work in progress), November 2020.

[I-D.ietf-bier-ping]

Nainar, N., Pignataro, C., Akiya, N., Zheng, L., Chen, M., and G. Mirsky, "BIER Ping and Trace", draft-ietf-bier-ping-07 (work in progress), May 2020.

[I-D.ietf-bier-pmmm-oam]

Mirsky, G., Zheng, L., Chen, M., and G. Fioccola, "Performance Measurement (PM) with Marking Method in Bit Index Explicit Replication (BIER) Layer", draft-ietf-bier-pmmm-oam-09 (work in progress), December 2020.

[I-D.ietf-bier-prefix-redistribute]

Zhang, Z., Bo, W., Zhang, Z., Wijnands, I., and Y. Liu,
"BIER Prefix Redistribute", draft-ietf-bier-prefix-
redistribute-00 (work in progress), August 2020.

[I-D.xzlnp-bier-ioam]

Min, X., Zhang, Z., Liu, Y., Nainar, N., and C. Pignataro,
"Bit Index Explicit Replication (BIER) Encapsulation for
In-situ OAM (IOAM) Data", draft-xzlnp-bier-ioam-01 (work
in progress), January 2021.

[I-D.zhang-bier-babel-extensions]

Zhang, Z. and T. Przygienda, "BIER in BABEL", draft-zhang-
bier-babel-extensions-04 (work in progress), November
2020.

[I-D.zzhang-tsvwg-generic-transport-functions]

Zhang, Z., Bonica, R., and K. Kompella, "Generic Transport
Functions", draft-zzhang-tsvwg-generic-transport-
functions-00 (work in progress), November 2020.

[RFC7368] Chown, T., Ed., Arkko, J., Brandt, A., Troan, O., and J.
Weil, "IPv6 Home Networking Architecture Principles",
RFC 7368, DOI 10.17487/RFC7368, October 2014,
<<https://www.rfc-editor.org/info/rfc7368>>.

Authors' Addresses

Zheng(Sandy) Zhang
ZTE Corporation

EMail: zhang.zheng@zte.com.cn

Zhaohui Zhang (editor)
Juniper Networks

EMail: zzhang@juniper.net

IJsbrand Wijnands
Individual

EMail: ice@braindump.be

Mankamana Mishra
Cisco Systems

EMail: mankamis@cisco.com

Hooman Bidgoli
Nokia

EMail: hooman.bidgoli@nokia.com

Gyan Mishra (editor)
Verizon

EMail: gyan.s.mishra@verizon.com

BIER
Internet-Draft
Intended status: Standards Track
Expires: April 17, 2022

Z. Zhang
Juniper Networks
E. Rosen
Individual
D. Awduche
Verizon
L. Geng
China Mobile
G. Shepherd
Individual
October 14, 2021

Multicast/BIER As A Service
draft-zzhang-bier-multicast-as-a-service-03

Abstract

This document describes a framework for providing multicast as a service via Bit Index Explicit Replication (BIER) [RFC7279], and specifies a few enhancements to [draft-ietf-bier-idr-extensions] [RFC8279] [RFC8401] [RFC8444] to enable multicast/BIER as a service.

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 April 17, 2022.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Terminologies	3
1.2. A CDN of A Single Provider	4
1.2.1. IGP/BGP Interworking	5
1.3. A CDN That Involves Another Provider	6
1.3.1. Providing Independent BAAS To Multiple Customers . .	6
1.3.2. Control and Accounting	7
1.4. Sets and Segmentation	8
1.4.1. Multiple Sets	8
1.4.2. Segmentation	8
2. Specifications for Enhancements to BIER Signaling with BGP/IGP	9
2.1. BGP Procedures	9
2.2. ISIS/OSPF Procedures	10
3. IANA Considerations	11
4. Security Considerations	11
5. Contributors	11
6. Acknowledgements	12
7. References	12
7.1. Normative References	12
7.2. Informative References	13
Authors' Addresses	13

1. Introduction

Currently multicast is primarily used in the following scenarios:

- o Enterprise Applications. For example, large scale financial data publishing.
- o Provider/underlay tunnels for MVPN and for EVPN BUM.

- o Real-time IPTV offered by a service provider to its customers.

Besides the above, large scale multicast services, especially transit multicast transport provided by large Internet Service Providers is virtually non-existent. This is mainly because of the following chicken and egg dilemma:

- o Traditional multicast technologies are complicated and lack scalability. The revenue that multicast services bring in cannot offset the Capex and Opex that an operator has to invest, so provider networks typically do not enable multicast even though the deployed equipment does support multicast.
- o As a result, Content Providers cannot take advantage of multicast and instead use less efficient methods like Ingress Replication, Peer2Peer, or multicast at application layer.

A recent multicast technology breakthrough, BIER, provides a simple and scalable solution for large scale multicast deployment, independent of number of multicast flows. In the meantime, large scale distribution of ultra high definition video content has become more and more popular and important. Service providers simply cannot keep on increasing their network capacity even if they could shift cost to Content Providers. With these developments, service providers now have both the need and means to provide scalable multicast service, potentially across multiple providers.

This document describes a framework for Multicast As A Service (MAAS) enabled by BIER. We use Content Delivery Network (CDN) as example, though it applies to any large scale multicast delivery service.

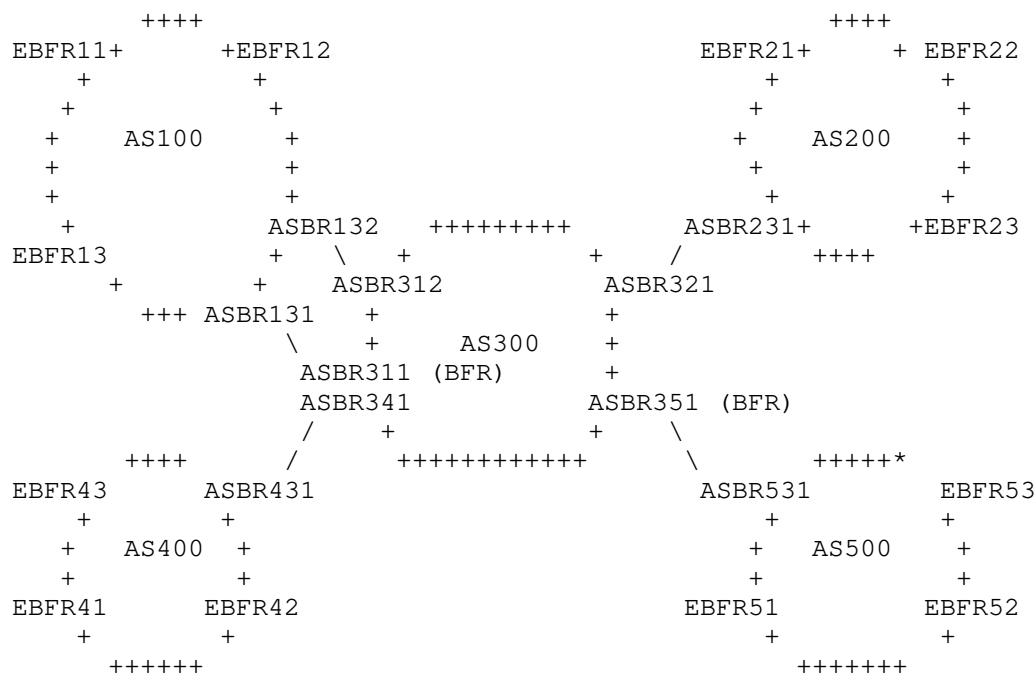
1.1. Terminologies

Readers are assumed to be familiar with multicast, BIER, BGP and ISIS/OSPF concepts and procedures. Some terminologies are listed here for convenience.

- o BFR: BIER Forwarding Router.
- o BFIR: BIER Forwarding Ingress Router.
- o BFER: BIER Forwarding Egress Router.
- o EBFR: Edge BFR. Including BFIR and BFER.
- o BSL: BitStrengLength. Number of bits in the BitString of a BIER header.

1.2. A CDN of A Single Provider

To make it easier to understand, we first consider a simple example: a CDN owned by a single operator, which could be a Content Provider itself. The network spans multiple ASes as shown in the following figure:



The CDN uses BIER for multicast transport and Edge BIER Forwarding Routers (EBFRs) are located throughout the network. Some of them are connected towards multicast content sources and are referred to as BIER Forwarding Ingress Routers (BFIRs) in BIER architecture. Most of them are connected towards multicast content receivers and are referred to as BIER Forwarding Egress Routers (BFERs). Notice that between content sources and BFIRs there may be Protocol Independent Multicast (PIM) in use, while between content receivers and BFERs there may be PIM and/or IGMP in use.

At the initial deployment stage, there might be only a few transit BIER Forwarding Routers (BFRs) at strategic points in the network (e.g. ASBR311 and ASBR351). BGP sessions are established among the EBFRs and BFRs, and BGP extensions as defined in

[I-D.ietf-bier-idr-extensions] are used to signal BIER information. All these are in a single BIER sub-domain.

In the example of initial stage with only ASBR311 and ASBR351 as BFRs, multicast traffic arriving at EBFR11 will be imposed with a BIER header and replicated to EBFR12/EBFR13/ASBR311 over tunnels. ASBR311 will further replicate traffic to ASBR351/EBFR41/EBFR42/EBFR43/EBFR21/EBFR22/EBFR23 over tunnels, and ASBR351 will further replicate traffic to EBFR51/EBFR52/EBFR53 over tunnels.

The BGP signaling and a necessary enhancement can be explained using the following example. EBFR43 advertises its BIER prefix (a loopback address) as /32 IPv4 or /128 IPv6 prefix in BGP with a BIER Path Attribute (BPA) [RFC8279] [I-D.ietf-bier-idr-extensions]. ASBR431 receives it and re-advertises it (with BGP Next Hop changed to itself) but does not do anything wrt BIER because it does not support BIER. Same happens on ASBR341. When ASBR311 and ASBR351 receive it from ASBR431, they create a BIFT entry corresponding to EBFR43's BFR-ID. The entry causes a BIER packet with corresponding bit set in its BitString to be tunneled to EBFR43. This cannot be based on BGP Next Hop in the advertisement because the BGP Next Hop is ASBR431. When eventually EBFR11 receives the re-advertised route, it creates a BIFT entry that causes corresponding packets to be tunneled to ASBR311 (but not to EBFR43 directly). Now it is clear that this cannot be based on either the BIER prefix itself or the BGP Next Hop. The solution is that the originating EBFR attaches a Tunnel Encap Attribute (TEA) [RFC9012] with the tunnel destination set to itself, and whenever a BFR re-advertises the route it changes the tunnel destination to itself. When a BFR creates the BIFT entry, it uses the Tunnel Egress Endpoint in the TEA to find out where to tunnel packets.

Over time, more routers in network may be upgraded to support BIER and become a BFR. For example, once ASBR431 is upgraded to a BFR, ASBR311 no longer needs to tunnel traffic to EBFR41/EBFR42/EBFR43 but only need to tunnel one copy to ASBR431, who will then replicate to EBFR41/EBFR42/EBFR43.

1.2.1. IGP/BGP Interworking

Additionally, if enough routers in an AS (or just one of its IGP areas) can be upgraded to run BIER, then hop-by-hop BIER forwarding can be utilized there, using IGP extensions for BIER signaling [RFC8401] [RFC8444].

Notice that even with this there is still only one BIER sub-domain, with mixed IGP and BGP signaling for BIER. To redistribute BIER

information between IGP and BGP, procedures specified in [I-D.ietf-bier-prefix-redistribute] and detailed in Section 2.2 are followed.

1.3. A CDN That Involves Another Provider

In the above example, the CDN is providing multicast transport service, with simplicity and scalability provided by BIER (the per-flow state is confined to the edges). Now let us go one step further and consider that AS300 belongs to a different Internet Service Provider. Now the ISP is providing BIER As A Service (BAAS) to the CDN, by being part of the CDN's BIER sub-domain. Notice that, not only does the ISP not have per-tree state (it does not have EBFRs), but also its BFRs do not need BFR-ID assigned. The ISP does need to learn about all the EBFRs and their corresponding BFR-IDs (through signaling).

1.3.1. Providing Independent BAAS To Multiple Customers

Now consider that the ISP also provides BAAS for another CDN. Each of the two CDNs has its own BIER domain, with its own BFR-ID or even sub-domain ID assignment that could conflict between the two CDNs. For example, both have BFR-ID 100 and sub-domain ID 0 assigned but they are totally independent of each other. For an BFR in the ISP to support this, with BGP signaling it needs to advertise its own BFR prefix multiple times, each time with a different RD that is mapped to the corresponding CDN. A new SAFI BIER (to be allocated by IANA) is used.

In the above example, there are two paths between AS100 and AS300. It is possible that while ASBR311 is the BFR, ASBR312 is the unicast best path into AS300 and beyond from AS100. Advertising BIER prefixes using a different SAFI with a RD also has the side benefit of allowing incongruent topologies for unicast and BIER.

In the existing BIER architecture and IGP extensions for BIER a sub-domain is tied to a single topology (either the one and only topology if Multi-topology ISIS/OSPF is not used, or a topology as defined in Multi-topology ISIS/OSPF). In the BIER sub-TLV that ISIS/OSPF attaches to a BIER prefix, a Sub-domain-ID value can only appear once for a particular topology. In this document, a BFR in the BAAS provider may belong to different and independent BIER domains, and the same sub-domain ID needs to be signaled multiple times, once for each BIER domain (notice that the same sub-domain-ID actually identifies different sub-domains in different BIER domains, so this does not really change the architectural requirement that a sub-domain is tied to a single topology). To do so, a new "BIER Domain" sub-TLV is introduced, and its value field includes a RD (as in the

BGP signaling) and a BIER sub-sub-TLV that is the same as currently specified in ISIS/OSPF extensions for BIER.

This works very well because of the flexible BIER architecture - a BIER packet is forwarded based on a Bit Index Forwarding Table (BIFT) that is determined by a 20-bit BIFT ID in front of the BIER header, and each (subdomain, BSL, set) tuple has its own BIFT. Traditionally, a subdomain is identified by a sub-domain ID but in this document a subdomain is now identified by a (RD, sub-domain ID) tuple in the control plane.

With this, the scaling aspect on a BFR comes to how many BAAS customer the provider needs to support. For example, if it needs to support 16 BAAS customers, one BSL, and four sets (Section 1.4.1) for each customer, then the provider needs to support 64 BIFTs ($16 \times 1 \times 4$). If the BSL is 256, then each BIFT has 256 entries in it and the total number of BIFT entries (routes) is 4k (256×64). Notice that this 4k number is not related to the number of customers' multicast flows, but only related to the number of customers and number of customer EBFRs. The number of customers with their own independent BIER domains are likely not very large initially, but if multicast as a service gets more widely used, the protocol and procedures defined in this document can scale up to the extent of how many BIFTs (and BIFT entries) a BFR can support. Since there is no real difference between a BIFT entry and a unicast RIB/FIB entry, as long as the scaling requirements are adequately considered in the BIER forwarding plane implementation (e.g., enough memory is allocated for the BIFTs), scaling will not become a bottleneck.

Building/updating the BIFTs is the same as in the base BIER architecture, except that in the control plane a subdomain is identified by a (RD, sub-domain ID) tuple instead of just a sub-domain ID. This is transparent to the forwarding plane - a BIFT is always identified by an opaque 20-bit opaque number. This opaque number is either a label for MPLS encapsulation or an opaque number for non-MPLS encapsulation, and the optional static encoding as specified in [I-D.ietf-bier-non-mpls-bift-encoding] cannot be used.

1.3.2. Control and Accounting

With BGP based signaling, internal routers of a BAAS provider does not need explicit configuration for the BIER transport services that it support. In the above example, the ASBRs (ASBR311, ASBR312, ASBR321, ASBR341, ASBR351) in AS300 only need to have BGP policy configured to allow certain received BIER prefix advertisements to trigger necessary BIER state and additional signaling of their own. For example, when ASBR351 receives the BIER prefix advertisement, if its local configuration allows it may create corresponding BIFTs and

BIFT entries, and additionally originates or updates its own BIER prefix advertisement. An internal BFR inside AS300, upon receiving the BGP advertisements, may or may not need to go through the same policy check again (based on the providers operation model).

When the ASBRs (re-)advertise BIER prefixes toward their external peers, they could enable statistics counters for the corresponding BIER labels so that they can count incoming BIER packets from external peers specifically for this BAAS. Similarly, the ASBRs can enable statistics counters for BIER labels they receive from external peers, so that they can count outgoing BIER packets delivered to the external peers. These incoming and outgoing counters can be used for accounting and billing purposes.

1.4. Sets and Segmentation

The number of EBFRs could very well be larger than the BSL. There are two ways to handle that - multiple sets or segmentation.

1.4.1. Multiple Sets

With this method the set of EBFRs are grouped into multiple sets, and the number of EBFRs in a set is smaller than the BSL. A BFIR may need to send multiple copies of a multicast packet to reach all BFRs, one copy for each set that covers one or more expecting BFRs. A separate BIFT is needed for each set (because the same bit in the BitString of packets for different sets maps to different BFRs). This not only leads to multiple copies to be sent over the same link, but also requires additional BIFTs. In the earlier example, 64 BIFTs are needed for 16 BAAS customers because each customer needs 4 BIFTs for the multiple sets.

1.4.2. Segmentation

With this method, a BIER network is segmented into multiple regions, each with its own BIER sub-domain. In the earlier example, each AS could be an independent sub-domain. A BIER packet from EBFR11 will be decapsulated by the segmentation border router ASBR311, and then sent into next sub-domain in AS300 with a new BIER header. The segmentation [RFC7524] involves Multicast Flow Overlay [RFC8279] [RFC8556] so that the segmentation border routers know what BitString to use when sending onto the next segment. The advantage of segmentation is that only a single copy needs to be sent, and the number of BIFTs is also reduced on all BFRs. The disadvantage is that the segmentation points need to run multicast flow overlay protocol and maintain related state in control plane and data plane.

A deployment may start without the need for either multiple sets or segmentation when the number of EBFRs is small. When the number of EBFRs grows, segmentation can be introduced incrementally. A new BFR can be added as, or an existing BFR could be converted to, a segmentation point, splitting the original sub-domain into two independent sub-domains. The segmentation point does not re-advertise BIER information from one sub-domain to another. Other BFRs/EBFRs do not need any configuration changes except to make sure that all BIER information exchange is restricted to a single sub-domain (for example, two BFRs were BGP peers before and were exchanging BIER information but now they belong to two sub-domains and only exchange BIER information with the segmentation point and other BFRs in the same sub-domain).

In the earlier example of a CDN of a single provider, using segmentation may be acceptable, even though the overlay state needs to be kept by the segmentation points. A BAAS provider may need to carefully consider if it wants to keep a customer's overlay state on those segmentation points. On the other hand, the provider may consider hosting per-customer segmentation points. For example, tethering small or virtual BFRs to an ASBR and have those BFRs be the segmentation points [I-D.ietf-bier-tether].

2. Specifications for Enhancements to BIER Signaling with BGP/IGP

2.1. BGP Procedures

When an EBFR advertises a BIER prefix with a BIER Path Attribute (BPA), it SHOULD attach a Tunnel Encap Attribute (TEA) with the tunnel destination set to itself.

A BFR receiving the advertisement MUST use the tunnel destination in the TEA to determine where to forward a BIER packet whose BitString has a set bit corresponding to the BIER prefix, unless the TEA does not exist, in which case the BIER prefix itself is used for the determination. When the BFR re-advertises the BIER prefixes, it MUST change the tunnel destination in the TEA to itself, or add a TEA with the tunnel destination set to itself if there was no TEA in the received advertisement.

The TEA SHOULD have a Protocol Sub-TLV with protocol type BIER (0xAB37).

A transit BFR that is allowed (by provisioning or based on policy) to participate in a BIER sub-domain MUST advertise its own BIER prefix with a BPA. The BFR-id in the BPA SHOULD be 0. Depending on the operational model of the operator, the advertisement MAY be based on

received BIER prefixes (subject to certain BGP policy verification), or MAY do so only with explicit configuration.

If a provider provides independent BAAS services to multiple customers, when its BFR receives BIER prefixes from a customer it MUST re-advertise with a new BIER SAFI. For simplicity, all BFRs of the provider use the same RD that is specifically assigned for the customer. When a BFR re-advertises BIER prefixes to a customer, it MUST re-advertise with SAFI 1 or 2.

If multiple providers together provide BAAS to a customer, then the two providers may assign the same RD for the customer or do RD rewriting when re-advertising BIER prefixes from one provider to another.

2.2. ISIS/OSPF Procedures

This document defines a new BIER Domain Sub-TLV of ISIS TLVs 135, 235, 236, and 237. The sub-TLV type is to be allocated.

This document also defines a new BIER Domain Sub-TLV of OSPF Extended Prefix TLV. The sub-TLV type is to be allocated.

The value part of the BIER Domain Sub-TLV includes a 64-bit Route Distinguisher followed by one or more BIER Info Sub-TLV (as defined in [RFC8401] and [RFC8444] respectively) as its sub-sub-TLVs .

When a BFR redistribute a BIER prefix from BGP into ISIS/OSPF, if the BGP advertisement is of BIER SAFI, a BIER Domain sub-TLV is attached, with the RD part of the sub-TLV copied from the BGP advertisement. For each BIER TLV in the BPA, a BIER Info sub-sub-TLV is added in the BIER Domain sub-TLV, with the subdomain-id and BFR-id copied from the corresponding BIER TLV in the BPA, and the Encapsulation sub-sub-sub-TLV omitted because it is not needed.

If the BGP advertisement is of SAFI 1 or 2, BIER Info Sub-TLVs are constructed as above directly, without using a BIER Domain sub-TLV.

When a BFR redistribute a BIER prefix from ISIS/OSPF into BGP, if there is a BIER Domain sub-TLV in the corresponding ISIS LSP or OSPF LSA, the BGP advertisement is of BIER SAFI and the RD part of the NLRI is set to the RD from the BIER Domain sub-TLV. For each BIER Info sub-sub-TLV in the BIER Domain sub-TLV, a BIER TLV is included in the BPA, with the subdomain-id and BFR-id copied from the corresponding BIER Info sub-sub-TLV. The MPLS Encapsulation sub-TLV is omitted. The tunnel destination in the TEA is set to the BFR's BIER prefix.

If there is no BIER Domain sub-TLV in the corresponding ISIS LSP or OSPF LSA for the BIER Prefix, the BGP advertisement is of SAFI 1 or 2, and the BPA is constructed similar to the above (the only difference is that in this case BIER Info sub-TLVs are not part of a BIER Domain sub-TLV).

3. IANA Considerations

This document requests the following IANA assignments:

- o A sub-TLV type for BIER Domain Sub-TLV from ISIS "Sub-TLVs for TLVs 135, 235, 236, and 237" registry.
- o A sub-TLV type for BIER Domain Sub-TLV from OSPFv2 Extended Prefix Sub-TLV registry.
- o A BIER SAFI from Subsequent Address Family Identifiers (SAFI) registry.

4. Security Considerations

There are no security concerns wrt exchange of BIER information besides what have been discussed in [I-D.ietf-bier-idr-extensions] and [RFC8401] [RFC8444].

The tunnels between BFRs that are not directly connected are ideally auto-configured to reduce provisioning burdens. Given that they may span multiple ASes and MPLS may not always be available, BIER over UDP/GRE/IPv4/IPv6 becomes very convenient, though that has the same security concerns well discussed in "Security Considerations" of [RFC4023] and [RFC7510].

As one mitigation when the tunnel is not secured, a BFR MAY use source address filtering based on pre-provisioned or dynamically learned allowable addresses. With dynamic learning, if a BFR receives a BIER prefix with a BPA and a TEA (see Section 2.1), it sets up a forwarding filter to allow IP/GRE/UDP tunneling from the address encoded in the "Tunnel Egress Endpoint" sub-TLV of Tunnel TLVs in the TEA. While that is the address for this BFR to tunnel traffic to, this BFR will also likely receive tunneled traffic from that address.

5. Contributors

The following people also contributed to this document.

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

Gyan Mishra
Verizon
Email: hayabusagsm@gmail.com

6. Acknowledgements

The authors thank Lenny Giuliano and Antoni Przygienda for their review and suggestions.

7. References

7.1. Normative References

- [I-D.ietf-bier-idr-extensions]
Xu, X., Chen, M., Patel, K., Wijnands, I., and A. Przygienda, "BGP Extensions for BIER", draft-ietf-bier-idr-extensions-07 (work in progress), September 2019.
- [I-D.ietf-bier-prefix-redistribute]
Zhang, Z., Wu, B., Zhang, Z., Wijnands, I., and Y. Liu, "BIER Prefix Redistribute", draft-ietf-bier-prefix-redistribute-00 (work in progress), August 2020.
- [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>>.
- [RFC8401] Ginsberg, L., Ed., Przygienda, T., Aldrin, S., and Z. Zhang, "Bit Index Explicit Replication (BIER) Support via IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018, <<https://www.rfc-editor.org/info/rfc8401>>.
- [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>>.
- [RFC8556] Rosen, E., Ed., Sivakumar, M., Przygienda, T., Aldrin, S., and A. Dolganow, "Multicast VPN Using Bit Index Explicit Replication (BIER)", RFC 8556, DOI 10.17487/RFC8556, April 2019, <<https://www.rfc-editor.org/info/rfc8556>>.

- [RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <<https://www.rfc-editor.org/info/rfc9012>>.

7.2. Informative References

- [I-D.ietf-bier-non-mpls-bift-encoding]
Wijnands, I., Mishra, M., Xu, X., and H. Bidgoli, "An Optional Encoding of the BIFT-id Field in the non-MPLS BIER Encapsulation", draft-ietf-bier-non-mpls-bift-encoding-04 (work in progress), May 2021.
- [I-D.ietf-bier-tether]
Zhang, Z., Warnke, N., Wijnands, I., and D. Awduche, "Tethering A BIER Router To A BIER incapable Router", draft-ietf-bier-tether-01 (work in progress), January 2021.
- [RFC7524] Rekhter, Y., Rosen, E., Aggarwal, R., Morin, T., Grosclaude, I., Leymann, N., and S. Saad, "Inter-Area Point-to-Multipoint (P2MP) Segmented Label Switched Paths (LSPs)", RFC 7524, DOI 10.17487/RFC7524, May 2015, <<https://www.rfc-editor.org/info/rfc7524>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks

EMail: zzhang@juniper.net

Eric Rosen
Individual

EMail: erosen52@gmail.com

Daniel Awduche
Verizon

EMail: daniel.awduche@verizon.com

Liang Geng
China Mobile

EMail: gengliang@chinamobile.com

Greg Shepherd
Individual

EMail: gjshep@gmail.com

tswwg
Internet-Draft
Intended status: Standards Track
Expires: May 5, 2021

Z. Zhang
R. Bonica
K. Kompella
Juniper Networks
November 01, 2020

Generic Transport Functions
draft-zzhang-tswwg-generic-transport-functions-00

Abstract

Some functionalities (e.g. fragmentation/reassembly and Encapsulating Security Payload) provided by IPv6 can be viewed as independent of IPv6 or even IP entirely. This document proposes to provide those functionalities at different layers (e.g., MPLS, BIER or even Ethernet) independent of IP.

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 May 5, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Specifications	4
2.1. Generic Fragmentation Header	4
2.2. MPLS Signaling	5
2.2.1. BGP Signaling	5
2.2.2. IGP Signaling	6
2.3. Generic ESP/Authentication Header	6
3. Security Considerations	6
4. IANA Considerations	6
5. Acknowledgements	7
6. References	7
6.1. Normative References	7
6.2. Informative References	8
Authors' Addresses	8

1. Introduction

Consider an operator providing Ethernet services such as pseudowires, VPLS or EVPN. The Ethernet frames that a Provider Edge (PE) device receives from a Customer Edge (CE) device may have a larger size than the PE-PE path MTU (pMTU) in the provider network. This could be because

1. the provider network is built upon virtual connections (e.g. pseudowires) provided by another infrastructure provider, or
2. the customer network uses jumbo frames while the provider network does not, or
3. the provider-side overhead for transporting customers packets across the network pushes past the pMTU.

In any case, the provider simply cannot require its customers to change their MTU.

To get those large frames across the provider network, currently the only workaround is to encapsulate the frames in IP (with or without GRE) and then fragment the IP packets. Even if MPLS is used for service delimiting, IP is used for transportation (MPLS over IP/GRE). This may not be desirable in certain deployment scenarios, where MPLS is the preferred transport or IP encapsulation overhead is deemed excessive.

IPv6 fragmentation and reassembly are based on the IPv6 Fragmentation header below [RFC8200]:

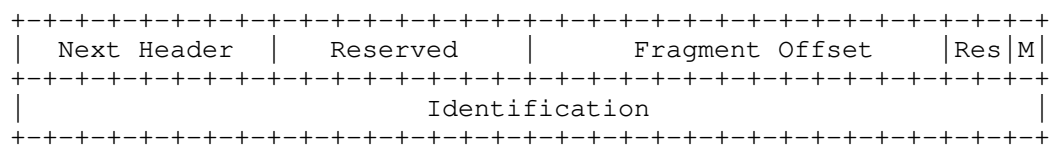


Figure 1: IPv6 Fragmentation Header

This document proposes reusing this header in non-IP contexts, since the fragmentation/reassembly function is actually independent of IPv6 except the following aspects:

- o The fragment header is identified as such by the "previous" header.
- o The "Next Header" value is from the "Internet Protocol Numbers" registry.
- o The "Identification" value is unique in the (source, destination) context provided by the IPv6 header

The "Identification" field, in conjunction with the IPv6 source and destination identifies fragments of the original packet, for the purpose of reassembly.

Therefore, the fragmentation/reassembly function can be applied at other layers as long as a) the fragment header is identified as such; and b) the context for packet identification is provided. Examples of such layers include MPLS, BIER, and Ethernet (if IEEE determines it is so desired).

For the layers where the IETF is concerned, the "Next Header" value will still be from the "Internet Protocol Numbers" registry when the function is applied at non-IP layers.

For the same consideration, the IP Encapsulating Security Payload (ESP) [RFC4303] could also be applied at other layers if ESP is desired there. For example, if for whatever reason the Ethernet service provider wants to provide ESP between its PEs, it could do so without requiring IP encapsulation if ESP is applied at non-IP layers.

The possibility of applying some other IP functions (e.g. Authentication Header [RFC4302]) is for further study.

2. Specifications

2.1. Generic Fragmentation Header

For generic fragmentation/reassembly functionality independent of IP, the following Generic Fragmentation Header (GFH) is defined:

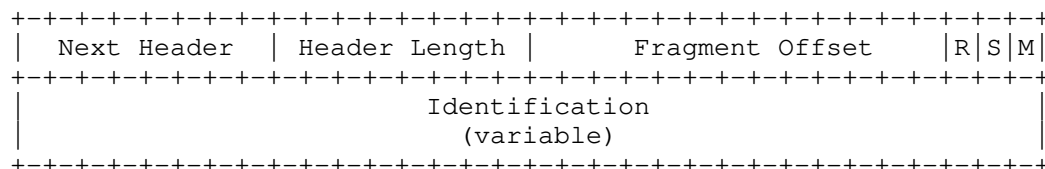


Figure 2: Generic Fragmentation Header

The "Next Header", "Fragment Offset" and "M" flag bit fields are as in the IPv6 Fragmentation Header.

Header Length: the number of octets of the entire header.

R: The "R" flag bit is reserved. It MUST be 0 on transmitting and ignored on receiving.

Identification: at least 4-octet long.

S: If the "S" flag bit is clear, the context for the Identification field is provided by the outer header, and only the source-identifying information in the outer header is used. If the "S" flag bit is set, the variable Identification field encodes both source-identifying information (e.g. the IP address of the node adding the GFH) and an identification number unique within that source.

The outer header MUST identify that a Generic Fragmentation Header follows and MAY carry source-identifying information.

If the outer header is BIER, a TBD value for the "proto" field in the BIER header identifies that a GFH follows. If the "S" flag bit is clear, the "BFIR-id" field in the BIER header provides the context for the "Identification" field.

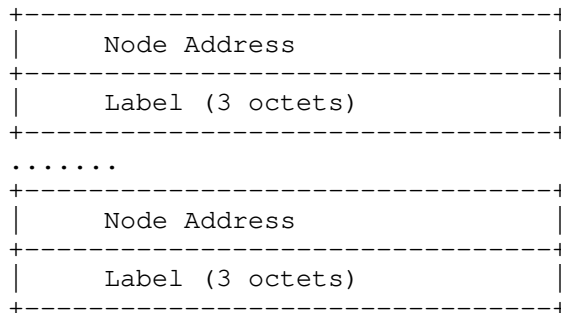
If the outer header is MPLS, the "S" flag bit MAY be clear if the the label preceeding the GFH identifies the sending BFR in addition to indicating that a GFH follows (see Section 2.2).

2.2. MPLS Signaling

When GFH is used with MPLS, the preceeding label needs to indicate that a GFH follows, and optionally identify the node that does the fragmentation. The label can be signaled via BGP or IGP as sepcified below.

2.2.1. BGP Signaling

This document defines a new transitive BGP "GFH Labels" attribute, very similar to the "PE Distinguisher Labels" attribute defined in [RFC6514] (and the text below is adapted from Section 8 of [RFC6514]):



The Label field contains an MPLS label encoded as 3 octets, where the high-order 20 bits contain the label value. The Node Address MAY be 0, meaning that the following label only indicates a GFH follows when the label is used in the label stack of a data packet.

The Node Address MAY also be a unicast address, indicating that the following label when used in the label stack of a data packet will both indicate that a GFH follows and identify the sending node.

If a node supports GFH with MPLS, it attaches the attribute in the BGP routes for its local addresses. A border router SHOULD remove the attribute if no node beyond the border will use GFH with MPLS to send traffic to the corresponding addresses.

A router that supports the attribute considers this attribute to be malformed if the Node Address field does not contain a unicast address or 0. The attribute is also considered to be malformed if: (a) the Node Address field is expected to be an IPv4 address, and the length of the attribute is not a multiple of 7 or (b) the Node Address field is expected to be an IPv6 address, and the length of the attribute is not a multiple of 19. The Address Family Indicator (AFI) of the BGP route that the attribute is attached to provides the

information on whether the Node Address field contains an IPv4 or IPv6 address. Each of the Node Addresses in the attribute MUST be of the same address family as the route that is carrying the attribute.

2.2.2. IGP Signaling

This document defines an OSPFv2 "GFH Labels" sub-TLV of OSPFv2 Extended Prefix TLV [RFC7684], with the value part being the same as BGP "GFH Labels" attribute above. If an OSPFv2 router supports GFH with MPLS, it includes the GFH Labels sub-TLV in the Extended Prefix TLV that is attached to its local addresses advertised in its OSPFv2 Extended Prefix Opaque LSA.

Similarly, This document defines an OSPFv3 "GFH Labels" sub-TLV of OSPFv3 Intra/Inter-Area-Prefix TLVs [RFC8362], with the value part being the same as BGP "GFH Labels" attribute above. If an OSPFv3 router supports GFH with MPLS, it includes the GFH Labels sub-TLV in the Intra-Area-Prefix TLV for its local addresses.

This document also defines an ISIS "GFH Labels" sub-TLV of ISIS prefix-reachability TLV [RFC5120] [RFC5305] [RFC5308], with the value part being the same as BGP "GFH Labels" attribute above. If an ISIS router supports GFH with MPLS, it includes the sub-TLV to the prefix-reachability TLV for its local addresses.

For both OSPF and ISIS, when advertising a prefix from one area/level to another, if there is a "GFH Labels TLV" attached in the source area/level, the TLV SHOULD be attached in the target area/level and the prefix SHOULD NOT be summarized.

2.3. Generic ESP/Authentication Header

To be specified in future revisions.

3. Security Considerations

To be provided.

4. IANA Considerations

This document makes the following IANA requests:

- o A new BGP Attribute type for "GFH Labels" from the BGP Path Attributes registry
- o A new OSPFv2 sub-TLV type for "GFH Labels" from the OSPFv2 Extended Prefix TLV Sub-TLVs registry

- o A new OSPFv3 sub-TLV type for "GFH Labels" from the OSPFv3 Extended-LSA sub-TLV registry
- o A new BIER Next Protocol Identifier value for GFH from BIER Next Protocol Identifiers registry

5. Acknowledgements

6. References

6.1. Normative References

- [RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, DOI 10.17487/RFC4303, December 2005, <<https://www.rfc-editor.org/info/rfc4303>>.
- [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <<https://www.rfc-editor.org/info/rfc5120>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<https://www.rfc-editor.org/info/rfc5305>>.
- [RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, DOI 10.17487/RFC5308, October 2008, <<https://www.rfc-editor.org/info/rfc5308>>.
- [RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, DOI 10.17487/RFC7684, November 2015, <<https://www.rfc-editor.org/info/rfc7684>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.
- [RFC8362] Lindem, A., Roy, A., Goethals, D., Reddy Vallem, V., and F. Baker, "OSPFv3 Link State Advertisement (LSA) Extensibility", RFC 8362, DOI 10.17487/RFC8362, April 2018, <<https://www.rfc-editor.org/info/rfc8362>>.

6.2. Informative References

- [RFC4302] Kent, S., "IP Authentication Header", RFC 4302, DOI 10.17487/RFC4302, December 2005, <<https://www.rfc-editor.org/info/rfc4302>>.
- [RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <<https://www.rfc-editor.org/info/rfc6514>>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks
1133 Innovation Way
Sunnyvale 94089
USA

Phone: +1 408 745 2000
Email: zzhang@juniper.net

Ron Bonica
Juniper Networks
1133 Innovation Way
Sunnyvale 94089
USA

Phone: +1 408 745 2000
Email: rbonica@juniper.net

Kireeti Kompella
Juniper Networks
1133 Innovation Way
Sunnyvale 94089
USA

Phone: +1 408 745 2000
Email: kireeti@juniper.net