Workgroup: Network Working Group
Internet-Draft:
draft-chen-bier-egress-protect-04
Published: 22 December 2022
Intended Status: Standards Track
Expires: 25 June 2023
Authors: H. Chen M. McBride A. Wang
Futurewei Futurewei China Telecom
G. Mishra Y. Liu
Verizon Inc. China Mobile
M. Menth B. Khasanov X. Geng
University of Tuebingen Yandex LLC Huawei
Y. Fan L. Liu X. Liu
Casa Systems Fujitsu Volta Networks

BIER Egress Protection

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 is called BIER egress protection. It does not require any per-flow state in the core of the domain. With BIER egress protection the failure of a primary BFER (Bit Forwarding Egress Router) is protected with a backup BFER such that traffic destined to the primary BFER in the BIER domain is fast rerouted by a neighbor BFR to the backup BFER on the BIER layer. The mechanism is applicable if all BIER traffic sent to the primary BFER can reach its destination also via the backup BFER. It is complementary to BIER-FRR which cannot protect against the failure of a BFER.

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

Copyright Notice

Table of Contents

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.

Basic LFA: It is the LFA defined in [RFC5286].

RLFA: Remote LFA. It is the LFA defined in [RFC7490].

TI-LFA: Topology Independent LFA. It is the LFA defined in [I-D.ietf-rtgwg-segment-routing-ti-lfa].

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.

FIB: Forwarding Information Base or Forwarding Table.

2. Overview of BIER Egress Protection

This section introduces BIER egress protection and describes its operation using the BIER topology in Figure 1 as an example. The figure illustrates a BIER sub-domain with the 8 nodes/BFRs A, B, C, D, E, F, G and H. Each link connecting these nodes/BFRs has a cost. The cost of a link (for routing purposes) is indicated in the figure unless it is 1 by default. Nodes/BFRs D, F, E, H and A are BFERs 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 5 bits long. BFR-ids 1, 2, 3, 4, and 5 are represented by (0:00001), (0:00010), (0:00100), (0:01000) and (0:10000), respectively.

```
                                          (CE2) Receiver
                                            \
                                             \  4 (0:01000)
                      /--------( G )--------- ( H ) Backup Egress for D
                    2/           2\______      /    \
                    /               _____)____/      \
                   /               /     (____        (CE1) Receiver
                  /               /           \      /
                 /               /             \    /
( A )----------( B )-----------( C )----------( D ) Primary Egress
  5 (0:10000)    \               \              1 (0:00001)
                 4\               \
                   \               \
                  ( E )-----------( F )
                    3 (0:00100)     2 (0:00010)
```

Figure 1: Example BIER topology

CE1 and CE2 in neighboring networks are multicast traffic receivers. CE1 is connected to both BFER D and BFER H. CE2 is connected to H but it is not connected to D.

We explain BIER egress protection for primary BFER D using backup BFER H. At first, BFER H is configured to protect BFER D. In addition, whether primary egress D and backup egress H send their BIER packets' payloads to the same receiver CE1 (i.e., after decapsulating their BIER packets, whether they send the same decapsulated packets to the same receiver CE1) is configured. And then, this information is distributed to BFR D's neighbors (BFR C and BFR G) and the domain by IGP. BFR C, BFR G, and BFER H know that H is the backup egress to protect the primary egress D. Two different backup strategies or methods, Bit Protection Switching and Proxy Backup, are specified for two different configurations regarding to whether D and H send their BIER packets' payloads to the same receiver.

1. **Bit Protection Switching:** If a neighbor of D detects D's outage, it performs the following operations on all the packets that are destined to D. It clears the bit for destination D and sets the bit for H. Afterwards, these packets are forwarded towards H and eventually reach H which decapsulates them and delivers their payloads to the same receiver CE as D does.

2. **Proxy Backup:** If a neighbor as PLR of D detects D's outage, it reroutes a copy of the packet with D as a destination towards H. When H as backup BFER detects its primary BFER D's outage, H, acting as a proxy of D, decapsulates all the BIER packets with destination D and forwards their payloads according to D's forwarding behavior for the payloads.

Bit Protection Switching is well applicable to the case where primary egress D and backup egress H send their BIER packets' payloads to the same receiver CE1. In this case, after D decapsulates D's BIER packet (i.e., the BIER packet with BFER D as a destination), D sends the decapsulated packet (i.e., the payload of the BIER packet) to receiver CE1 through its multicast layer. After H decapsulates H's BIER packet (i.e., the BIER packet with BFER H as a destination), H sends the same decapsulated packet (i.e., the same payload as the one in D's BIER packet) to the same receiver CE1 through its multicast layer as D.

During normal operations, there is no multicast traffic to CE1 from backup egress H, and CE1 receives the multicast traffic only from primary egress D. There is no duplicated traffic to receiver CE1.

When primary egress D fails, the BIER packet with destination D is updated through bit switch (i.e., the bit for D is cleared and bit for H is set in the packet) by a PLR such as BFR C when the PLR detects the failure of D. The updated packet with destination H is sent to backup egress H. H decapsulates the packet and delivers the packet's payload to its multicast layer, which sends the payload to CE1.

Proxy Backup is applicable to the case where D and H send their BIER packets' payloads to different receivers. In this case, after D decapsulates D's BIER packet, D sends the decapsulated packet (i.e., the payload of the BIER packet) to receiver CE1 through its multicast layer. After H decapsulates H's BIER packet, H drops the same decapsulated packet (i.e., the same payload as the one in D's BIER packet) or sends it to different receiver CE2 through its multicast layer.

During normal operations, primary egress D sends the payload of the BIER packet with destination D to receiver CE1 and backup egress H sends the payload of the BIER packet with destination H to receiver CE2. H sends the BIER packet with destination D towards node D along the shortest path to D.

When D fails, the BIER packet with destination D is sent to backup egress H by a PLR such as BFR C when the PLR detects the failure of D. H acting as a proxy of D MUST have a fast way to detect the failure of D and obtain the forwarding behavior of D for the payload of the BIER packet with destination D in advance. When H as the proxy of D detects the failure of D, it sends the payload of the BIER packet with destination D to receiver CE1 according to the forwarding behavior of D for the payload.

Backup egress H may obtain the forwarding behavior of its primary egress D for the payload of the BIER packet with the primary egress as a destination from configurations or through some protocols such as BGP or PCEP. How for a backup egress to obtain the forwarding behavior of its primary egress is out scope of this document.

The fast egress protection mechanism in this document is different from MoFRR in [RFC7431], where the same traffic is sent through two separated paths/trees to both primary egress node D and backup egress node H, to which the receiver CE1 is dual homed. It will use less network resources such as link bandwidth than MoFRR in [RFC7431].

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).

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 or 10. Using the LSA of LS type 9, node P will advertise the information only to its neighbors (which will not advertise the information further). Using the LSA of LS type 10, node P will advertise the information to the whole BIER network domain (i.e., P's neighbors will advertise the information further until the information reaches every node in the domain). The information is included in a backup egress node TLV. The format of the TLV is shown in Figure 2.

After each of the neighbors receives the backup egress node TLV, 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 BIER packet with the bit for destination P towards node P's backup egress node.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type (TBD1)           |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Reserved            |S| BFR-id of backup egress node  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Sub-TLVs (Optional)                     |
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: 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: 15 bits, they MUST be set to zero when sending and be ignored while receiving.

S flag: 1 bit. It is set to one to indicate that the primary egress and backup egress send their BIER packets' payloads to the same

CE receiver ; it is set to zero to indicate that the primary egress and backup egress send their BIER packets' payloads to different CE receivers .

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.

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 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]. Using CS-LSP or IIH PDUs, node P will advertise the information only to its neighbors. Using LSPs, node P will advertise the information to the whole BIER network domain. The format of the TLV is shown in Figure 3.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Type (TBD2)  |     Length    |          Reserved           |S|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| BFR-id of backup egress node  |    Sub-TLVs (Optional)       ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: IS-IS Backup Egress TLV

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

Length: 1 octet, its value is 4 plus the length of the Sub-TLVs included. If no Sub-TLV is included, its value is 4.

The other fields are the same as those in Figure 2.

4. Extensions to BIFT

This section specifies the BIFT extended for egress protection (EP-BIFT) on a BFR as a PLR and the BIFT extended on a backup egress node. In one option, the EP-BIFT is implemented in an Integrated one BIFT. In another, it is implemented in Multiple Backup BIFTs.

4.1. Integrated one BIFT

A BFR has an integrated BIFT for both normal operations and protections against the failure of each of its neighbor BFERs. That is that the normal BIFT on the BFR is extended to have a backup entry (or say sub-entry) for each of its neighbor BFERs.

4.1.1. EP-BIFT on BFR as PLR

To protect a primary egress node (e.g., BFER D in Figure 1), a BFR as the primary egress node's neighbor (e.g., BFR C in Figure 1) and a PLR has a backup entry in its BIFT extended for egress protection (EP-BIFT). The backup entry contains: Backup Entry Active (BEA), Same CE receiver (SC), Backup Egress BFER (BE-BFER), Backup F-BM (BF-BM) and Backup BFR-NBR (BBFR-NBR).

* BEA = 1 indicates that the Backup Entry for egress protection is active.

* SC = 1 indicates that both primary egress node and backup egress node send their BIER packets' payloads to the same receiver CE.

* BE-BFER is the BFR-id of the backup egress node for the primary egress node.

* BBFR-NBR is the backup BFR-NBR to the backup egress node (e.g., H in Figure 1). When SC = 1 (i.e., both primary egress node and backup egress node send their BIER packets' payloads to the same receiver CE), the BFR finds a basic, remote or topology independent (TI) LFA to the backup egress node and sets BBFR-NBR to the LFA. When SC = 0 (i.e., the primary egress node and its backup egress node send their BIER packets' payloads to different receiver CEs), the BFR obtains the value of BBFR-NBR in following steps. At first, the BFR finds a basic, remote or TI LFA to the backup egress node. And then the BFR checks if the LFA is the backup egress node or the backup egress node is on the shortest path from the LFA to the primary egress node without going through the primary egress node. If so, the LFA is used as the BBFR-NBR; otherwise (i.e., the LFA is not the backup egress node and the backup egress node is not on the shortest path from the LFA to the primary egress node without going through the primary egress node), the BBFR-NBR is set to the backup egress node through a tunnel to the backup egress node without going through

the primary egress node. This is to make sure that the BIER packet with the primary egress node as a destination reaches the backup egress node.

When primary egress node (e.g., BFER D in Figure 1) fails, the BFR as a PLR sets BEA in the entry for primary egress node to one after the BFR detects the failure. The BFR uses the backup entry with BEA = 1 to forward the BIER packet with primary egress node as a destination. The BFR forwards the packet to BBFR-NBR. Before forwarding the packet, the BFR checks whether SC equals to one in the entry. If SC = 1, the BFR as a PLR replaces the primary egress node as a destination with its backup egress node as a destination through clearing the bit for primary egress node (e.g., D) as a destination in the BIER packet and setting the bit for backup egress node (e.g., H) as a destination in the packet.

For example, the integrated BIFT (or say EP-BIFT) on BFR C in Figure 1 is shown in Figure 4.

BFR-id (SI:BitString)	F-BM	BFR-NBR	BEA	SC	BE-BFER	BF-BM	BBFR-NBR
1 (0:00001)	00001	D	0	1	H(01000)	01001	H
2 (0:00010)	00110	F	0	0	E(00100)	00010	E(TI-LFA)
3 (0:00100)	00110	F	0	0	F(00010)	00110	F
4 (0:01000)	01000	H	0	1	D(00001)	01001	D
5 (0:10000)	10000	B	0		0	NULL	NULL

Figure 4: Integrated BIFT on BFR C

BFR C in Figure 1 has three neighbor BFERs D, F and H with BFR-ids 1, 2 and 4 respectively. The backup entry for BFER D with BFR-id = 1 is the last five columns in the first row of Figure 4.

*BEA = 0 means that D is working well.

*SC = 1 means that the primary egress node D and backup egress node H send their BIER packets' payloads to the same CE receiver.

*BE-BFER = H means that H is the backup egress node for primary egress node D.

*BF-BM = 01001 is computed by ORing the bit of BFR-id with BFR-NBR = H and the bit of BFR-id with BBFR-NBR = H. BFR-id = 1 is with BBFR-NBR = H and BFR-id = 4 is with BFR-NBR = H.

*BBFR-NBR = H means that BFER H is the next hop on the shortest path to H without going D.

The backup entry for BFER F with BFR-id = 2 is the last five columns in the second row of Figure 4.

*BEA = 0 means that F is working well.

*SC = 0 means that the primary egress node F and backup egress node E send their BIER packets' payloads to different CE receivers.

*BE-BFER = E means that E is the backup egress node for primary egress node F.

*BF-BM = 00010 is computed by ORing the bit of BFR-id with BFR-NBR = E and the bit of BFR-id with BBFR-NBR = E. Since there is no BFR-id with BFR-NBR = E, BF-BM = 00010.

*BBFR-NBR = E (TI-LFA) means that B and E in Figure 1 are not on the shortest path to E without going F and TI-LFA tunnel is used to send primary egress node F's BIER packet to backup egress node E when F fails and BEA is set to one.

The backup entry for BFER H is similar to the one for BFER D. The backup entry for BFER E is similar to the one for BFER F.

4.1.2. EP-BIFT on Backup Egress

If a primary egress node (e.g., D in Figure 1) and its backup egress node (e.g., H in Figure 1) send their BIER packets' payloads to the same receiver CE (e.g., CE1 in Figure 1), then the forwarding entry for the primary egress node in the BIFT on the backup egress node keeps the same as normal.

For example, the integrated BIFT on backup egress node H in Figure 1 with SC = 1 is the same as H's normal BIFT, which is illustrated in Figure 5.

BFR-id (SI:BitString)	F-BM	BFR-NBR
1 (0:00001)	10111	C
2 (0:00010)	10111	C
3 (0:00100)	10111	C
4 (0:01000)	01000	H
5 (0:10000)	10111	C

Figure 5: Integrated BIFT on Backup Egress H with SC = 1

If the primary egress node and the backup egress node send their BIER packets' payloads to different receiver CEs, for example, D as a primary egress node sends its BIER packet's payload to CE1, H as the backup egress node for D sends its BIER packet's payload to CE2, then the forwarding entry for the primary egress node on the backup egress node is extended to contain a backup entry for primary egress node. The backup entry includes:

* Backup Entry Active (BEA), SC, BE-BFER, Backup F-BM (BF-BM). These have the same meanings as those in Section 4.1.1.

* Backup BFR-NBR or Pointer to FIB for Primary Egress (BBFR-NBR/P-FIB) is a pointer to the FIB for the primary egress node. Using this FIB, the backup egress node will forward the payload of the BIER packet with the primary egress node as a destination to the same CE receiver as the primary egress node.

BEA is set to one when the backup egress node detects the failure of the primary egress node. After detecting the failure and receiving the BIER packet with the bit for the primary egress node as a destination set to one, the backup egress node forwards the packet's payload to the primary egress node's CE receiver using the backup forwarding entry with BEA = 1.

For example, the integrated BIFT on backup egress node H in Figure 1 with SC = 0 is illustrated in Figure 6.

BFR-id (SI:BitString)	F-BM	BFR-NBR	BEA	SC	BE-BFER	BF-BM	BBFR-NBR /P-FIB
1 (0:00001)	10111	C	0	0	H(01000)	00001	P-FIB-4D
2 (0:00010)	10111	C	0	0			NULL
3 (0:00100)	10111	C	0	0			NULL
4 (0:01000)	01000	H	0	0			NULL
5 (0:10000)	10111	C	0				NULL

Figure 6: Integrated BIFT on Backup Egress H with SC = 0

In Figure 6, the backup entry for primary egress node D with BFR-id = 1 is the last five columns in the first row.

* BEA = 0 means that D is working well.

* SC = 0 means that the primary egress node D and backup egress node H send their BIER packets' payloads to different CE receivers.

* BE-BFER = H means that H is the backup egress node for primary egress node D.

* BF-BM = 00001 is computed by ORing the bit of BFR-id with BFR-NBR = P-FIB-4D and the bit of BFR-id with BBFR-NBR = P-FIB-4D. Since there is no BFR-id with BFR-NBR = P-FIB-4D, BF-BM = 00001.

* BBFR-NBR/P-FIB = P-FIB-4D is the pointer to the FIB for the primary egress node D. When D fails and BEA is set to one, backup egress node H for D acts as a proxy of D and sends D's BIER packet's payload to CE receiver CE1 using the FIB for D. Backup egress node H for D decapsulates the BIER packet with D as a destination and forwards the payload using the FIB for D after it detects the failure of D.

4.1.3. Updated Forwarding Procedure for Integrated BIFT

The forwarding procedure defined in [RFC8279] is updated/enhanced for integrated BIFT to consider the egress protection.

For a multicast packet with the BitString indicating a BFER as one of its destinations, the updated forwarding procedure on a BFR as a PLR sends the packet towards the backup egress node of the BFER if the BFER is protected. On the backup egress, the procedure sends the packet's payload to the BFER's CE receiver.

It checks whether BEA = 1 in the forwarding entry for the BFER. If BEA = 1, it determines whether the current node is backup egress node. On backup egress node, the procedure sends the packet's payload to the CE receiver. On the BFR as a PLR, the procedure sends the packet copy to BBFR-NBR. Before sending the packet copy, the procedure updates the packet copy by clearing the bit for primary egress node and setting the bit for backup egress node when primary egress node and backup egress node send their BIER packets' payload to the same CE receiver. The bits for the other destinations which are not through BBFR-NBR are cleared in the packet copy's BitString by ANDing the BitString with BF-BM. The original packet's BitString is updated to remove the bits for the destinations towards which the packet copy is sent through BBFR-NBR by ANDing the BitString with the INVERSE of BF-BM.

The updated forwarding procedure for integrated BFIT is described in Figure 7.

```
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 BEA == 1 { // Primary Egress fails
        IF (BBFR-NBR/P-FIB is Pointer to FIB) {// on Backup Egress
           Sends payload to CE using the FIB for primary egress;
        } ELSE {
           IF (SC == 1) {// on PLR and SC == 1
              clears bit k in Packet's BitString;//BFER k is PE-BFER
              sets bit j in Packet's BitString;  //BFER j is BE-BFER
           } // SC == 0, no updates to packet
           Copies Packet, updates the copy's BitString by ANDing it
           with BF-BM in the entry, sends updated copy to BBFR-NBR;
         }
         updates Packet's BitString by ANDing it with
         the INVERSE of BF-BM;
      } ELSE {
         Copies Packet, updates the copy's BitString by ANDing
         it with F-BM in the entry, sends updated copy to BFR-NBR;
         updates Packet's BitString by ANDing it with the INVERSE
         of the F-BM in the entry
      }
   }
}
```

Figure 7: Updated Forwarding Procedure for Integrated BIFT

4.2. Multiple Backup BIFTs

A BFR has a normal BIFT and multiple backup BIFTs for egress protection. For each of the BFR's neighbor BFERs, the BFR has a backup BIFT for the BFER, which considers the failure of the BFER. In normal operations, the BFR uses its normal BIFT to forward all the BIER packets. When the BFR detects the failure of the BFER, the BFR uses the backup BIFT for the BFER to forward all the BIER packets.

4.2.1. Multiple Backup BIFTs on BFR as PLR

A BFR as a PLR has a backup BIFT for a BFER that has the same structure as the normal BIFT except for a backup BFER (BE-BFER) for the BFER and same CE receiver (SC) flag indicating whether the BE-BFER and BFER send their BIER packets' payloads to the same CE receiver. In the entry for the BFER in the backup BIFT, the value of

BFR-NBR is the backup BFR-NBR (BBFR-NBR), which is computed in the same way as the BBFR-NBR is computed in Section 4.1.1.

For example, the backup BIFT for BFER D on BFR C in Figure 1 is shown in Figure 8. The backup BIFT for D considers BFER D's failure.

BFR-id (SI:BitString)	F-BM	BFR-NBR	SC	BE-BFER
1 (0:00001)	01001	H	1	H(01000)
2 (0:00010)	00110	F		
3 (0:00100)	00110	F		
4 (0:01000)	01001	H		
5 (0:10000)	10000	B		

Figure 8: BFR C's Backup BIFT for BFER D

In Figure 8, the entry for BFER D with BFR-id = 1 has its BFR-NBR with value of the BBFR-NBR (which is H) and contains SC = 1 and BE-BFER = H. BE-BFER = H means that BFER H is the backup egress node for primary egress node D. SC = 1 means that primary egress node D and backup egress node H send their BIER packets' payloads to the same CE receiver.

For the entry with BFR-NBR = X, its F-BM has the bit of the BFR-id in each entry with BFR-NBR = X. For example, the first entry with BFR-NBR = H, its F-BM in the first entry has the bit of BFR-id = 1 and BFR-id = 4 in the first entry and the fourth entry, which are with BFR-NBR = H.

When BFR C detects the failure of BFER D, it uses the backup BIFT for D to forwards all the BIER packets. For the packet with destination D (i.e., BitString = 00001), BFR C sends the packet to BFR-NBR H after clearing the bit for primary egress node D and setting the bit for backup egress node H since SC = 1. The packet received by H contains BitString = 01000 for destination H. After receiving the packet, BFER H sends the packet's payload to the same CE receiver CE1.

If SC = 0, BFR C sends the packet to BFR-NBR H without clearing the bit for D or setting the bit for H. After receiving the packet with destination D (i.e., BitString 00001) and detecting the failure of

D, BFER H as a proxy of D sends the packet's payload to primary egress node D's CE receiver CE1.

4.2.2. Multiple Backup BIFTs on Backup Egress

When a primary egress node and its backup egress node send their BIER packets' payloads to the same CE receiver, the backup BIFT for the primary egress node on the backup egress node is the same as the normal BIFT on the backup egress node. For example, the backup BIFT for primary egress node on backup egress node H in Figure 1 with SC = 1 is the same as H's normal BIFT, which is illustrated in Figure 5.

When a primary egress node and its backup egress node send their BIER packets' payloads to different CE receivers, the backup BIFT for the primary egress node on the backup egress node considers the failure of the primary egress node. The BFR-NBR/P-FIB in the entry for the primary egress node is the pointer to the FIB for the primary egress node which is used to forward the payload of the BIER packet with the primary egress node as a destination. For example, the backup BIFT for primary egress node D on backup egress node H in Figure 1 with SC = 0 is illustrated in Figure 9.

BFR-id (SI:BitString)	F-BM	BFR-NBR /P-FIB	SC	BE-BFER
1 (0:00001)	00001	P-FIB-4D	0	H(01000)
2 (0:00010)	00110	C		
3 (0:00100)	00110	C		
4 (0:01000)	01001	H		
5 (0:10000)	10000	C		

Figure 9: Backup Egress H's Backup BIFT for Egress D

In Figure 9, the entry for BFER D with BFR-id = 1 has its BFR-NBR/P-FIB = P-FIB-4D (the pointer to the FIB for primary egress node D) and contains BE-BFER = H and SC = 0. BE-BFER = H means that BFER H is the backup egress node for primary egress node D. SC = 0 means that primary egress node D and backup egress node H send their BIER packets' payloads to different CE receivers. Note that the last two columns can be removed since they are not used for forwarding.

When backup egress node H detects the failure of primary egress node D, node H uses the backup BIFT for egress D to forward all the BIER packets. For the packet with destination D (i.e., BitString = 00001), node H as a proxy of D sends the packet's payload to the CE1 (D's CE receiver) using the FIB for BFER D, which contains the forwarding behavior of primary egress node D for the payload of D's BIER packet.

4.2.3. Updated Forwarding Procedure for Multiple BIFTs

The updated forwarding procedure for multiple BIFTs is illustrated in Figure 10. This forwarding procedure is used with the normal BIFT on a BFR in normal operations. It is used with a backup BIFT for a primary egress node on a BFR as a PLR and on a backup egress node when the primary egress node fails.

On the backup egress node (i.e., BFR-NBR/P-FIB is a pointer to the FIB for the primary egress node), the procedure sends the payload of the packet with primary egress node/BFER as a destination to the BFER's CE receiver.

The forwarding procedure on a BFR as a PLR for each of multiple backup BIFTs is the same as the one defined in [RFC8279] except for sending the packet with primary egress node as a destination to the backup egress node of primary egress node. Before sending the packet to the backup egress node, the procedure updates the BitString in the packet by clearing the bit for the primary egress node and setting the bit for the backup egress node when SC = 1 (i.e., the primary egress node and backup egress node send their BIER packets' payloads to the same CE receiver).

```
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 (BFR-NBR/P-FIB is Pointer to FIB) {// on Backup Egress
         Sends payload using the FIB for the primary egress;
      } ELSE {
         IF (SC == 1) {// on PLR and SC == 1
            clears bit k in Packet's BitString;//BFER k is PE-BFER
            sets bit j in Packet's BitString;  //BFER j is BE-BFER
         } // SC == 0, no updates to packet
         Copies Packet, updates the copy's BitString by ANDing
         it with F-BM in the entry, sends updated copy to BFR-NBR;
      }
      updates Packet's BitString by ANDing it with the INVERSE
      of the F-BM in the entry
   }
}
```

Figure 10: Updated Forwarding Procedure for Multiple BIFTs

4.2.4. Switching between EP and Normal Forwarding

When multiple backup BIFTs are used, the multiple backup BIFTs are pre-computed and installed ready for activation when an egress node failure is detected. In normal operations, a BFR uses its normal BIFT to forward BIER packets. Once the BFR detects the failure of its BFR-NBR X as an egress, it activates (i.e., uses) the backup BIFT for X to forward BIER packets and de-activates (i.e., does not use) its normal BIFT. After activation of the backup 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 backup BIFT for X is de-activated.

From the new topology, the BFR computes/re-computes the backup BIRT for each BFR-NBR Y as an egress and the backup BIFT for Y is derived/re-derived from the backup BIRT for Y. The backup 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 using multiple backup BIFTs on a BFR in a BIER topology in Figure 1.

5.1. 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 1, 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 11.

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 11: 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 12.

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 12: Bit Index Forwarding Table on BFR C

5.2. Backup BIRTs and Backup 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) or say backup BIRTs.

For the BIER topology in Figure 1,

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 backup 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 backup BIRTs for its three BFR-NBRs (BFERs D, F and H) that are egress nodes respectively.

The backup BIRT for BEFR D built by BFR C based on the BIRT on BFR C (refer to Figure 11) is shown in Figure 13.

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

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

Figure 13: C's Backup BIRT for BFER D

In the backup BIRT for BFER D, the row that has Destination BFER == D is the 1st row. This row has the new backup information 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 BE-BFER = 0/NULL.

The 1st row in the EP-BIRT indicates that the packet with destination D will be sent to D's backup egress node H when D fails.

The 2nd row in the backup 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 backup 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 backup 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 backup BIRT indicates that the next hop BFR-NBR on the path to BFER A with BFR-id 5 is BFR B.

From this backup BIRT for BFER D on BFR C, an Egress Protection Bit Index Forwarding Table (EP-BIFT) or say backup BIFT for BFER D is derived. This backup BIFT for BFER D is shown in Figure 14.

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

The 2nd and 3rd rows in the backup BIRT have the same next hop BFR-NBR = E. The F-BM for each of these two rows in the backup 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)	SC	BE-BFER
1 (0:00001)	01001	H	1	H
2 (0:00010)	00110	F	0	0
3 (0:00100)	00110	F	0	0
4 (0:01000)	01001	H	0	0
5 (0:10000)	10000	B	0	0

Figure 14: C's Backup BIFT for BFER D

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

5.3. Forwarding using Backup 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/BFERs 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 normal 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 backup 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 backup 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 backup BIFT for BFER D. BE-BFER = H in the row indicates that BFER D is protected against the failure of D by backup BFER H. 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 since SC = 1. The BitString in Packet is 01110 now. BFR C copies, updates the BitString by ANDing it with F-BM (which is 01001) and sends the packet copy with BitString = 01000 to BFR-NBR H in the entry.

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 first row. The updated BitString = 00110, which is 01110 & ~F-BM in the row = 01110 & 10110 = 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 C gets the 2nd row (i.e., forwarding entry) in the backup BIFT for BFER D. 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 00110 & F-BM in the 2nd row = 00110 & 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 = 00000, which is 00110 & ~F-BM in the row = 00110 & 11001 = 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 sends the packet to E.

6. Security Considerations

TBD.

7. IANA Considerations

No requirements for IANA.

8. Acknowledgements

The authors would like to thank Jeffrey Zhang, Jingrong Xie 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., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-08, 21 January 2022, <https://www.ietf.org/archive/id/draft-ietf-rtgwg-segment-routing-ti-lfa-08.txt>.

[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", Work in Progress, Internet-Draft, draft-ietf-spring-segment-protection-sr-te-paths-03, 7 March 2022, <https://www.ietf.org/archive/id/draft-ietf-spring-segment-protection-sr-te-paths-03.txt>.

[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,
United States of America

Email: Huaimo.chen@futurewei.com

Mike McBride
Futurewei

Email: michael.mcbride@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
United States of America

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

Yisong Liu
China Mobile

Email: liuyisong@chinamobile.com

Michael Menth
University of Tuebingen

Email: menth@uni-tuebingen.de

Boris Khasanov
Yandex LLC
Moscow

Email: bhassanov@yahoo.com

Xuesong Geng
Huawei

Email: gengxuesong@huawei.com

Yanhe Fan
Casa Systems
United States of America

Email: yfan@casa-systems.com

Lei Liu
Fujitsu
United States of America

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks
McLean, VA
United States of America

Email: xufeng.liu.ietf@gmail.com