Workgroup: Network Working Group
Internet-Draft: draft-chen-bier-te-enc-02
Published: 7 March 2023
Intended Status: Standards Track
Expires: 8 September 2023
Authors: H. Chen M. McBride R. Chen
Futurewei Futurewei ZTE Corporation
G. Mishra A. Wang Y. Fan
Verizon Inc. China Telecom Casa Systems
L. Liu X. Liu
Fujitsu IBM Corporation

BIER-TE Encapsulation with Multiple BitStrings

Abstract

This document describes a "Bit Index Explicit Replication Traffic Engineering" (BIER-TE) header, two levels of Bit Index Forwarding Tables (BIFTs) and a forwarding procedure for efficiently processing BIER-TE packets with the header. For a multicast packet with an explicit point-to-multipoint (P2MP) path, which has multiple BitStrings, the packet with the header containing the BitStrings is replicated and forwarded statelessly along the path.

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 8 September 2023.

Table of Contents

1. Introduction

[RFC9262] introduces Bit Index Explicit Replication (BIER) Traffic/Tree Engineering (BIER-TE). It is an architecture for per-packet stateless explicit point to multipoint (P2MP) multicast path/tree.

A Bit-Forwarding Router (BFR) in a BIER-TE domain has a BIER-TE Bit Index Forwarding Tables (BIFT). A BIER-TE BIFT on a BFR comprises a forwarding entry for a BitPosition (BP) assigned to each of the adjacencies of the BFR. If the BP represents a forward connected adjacency, the forwarding entry for the BP forwards the multicast packet with the BP to the directly connected BFR neighbor of the adjacency. If the BP represents a BFER (i.e., egress node) or say a local decap adjacency, the forwarding entry for the BP decapsulates

the multicast packet with the BP and passes a copy of the payload of the packet to the packet's NextProto within the BFR.

[RFC8296] defines the BIER header with one BitString with Default BitStringLength value of 256. However, for a BIER-TE path from an ingress to multiple egresses (or say destinations), the bit positions representing the path may not be in one BitString. The existing BIER header does not work for the BIER-TE path with more than one BitString.

This document proposes a solution for a BIER-TE header to resolve this issue. The BIER-TE header can contain all the bit positions of a BIER-TE path. These bit positions are encoded in one or more BitStrings. The document presents an enhanced forwarding procedure for efficiently processing the BIER-TE header with multiple BitStrings.

1.1. Terminology

BIER: Bit Index Explicit Replication.

BIER-TE: BIER Traffic Engineering.

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.

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.

IGP: Interior Gateway Protocol.

LSDB: Link State DataBase.

OSPF: Open Shortest Path First.

IS-IS: Intermediate System to Intermediate System.

SI: Set Identifier.

BP:

Bit Position.

2. Example BIER-TE Path with Multiple BitStrings

This section illustrates an example BIER-TE domain topology and a BIER-TE paths across the domain. The path has multiple sets of bit strings, i.e., multiple BitStrings with different SIs (or multiple BitStrings for short). The packet to be transported by this path must contains the multiple BitStrings in the header of the packet. If the header can contain only one BitString, the packet to be transported by the path cannot be delivered to the egresses of the path.

2.1. Example BIER-TE Topology

An example BIER-TE topology for a BIER-TE domain is shown in Figure 1. It has 9 nodes/BFRs A, B, C, D, E, F, G, H and I. Nodes/BFRs D, F, E, H and A are BFERs and have local decap adjacency BitPositions 1, 2, 3, 4, and 5 respectively. For simplicity, these BPs are represented by (SI:BitString), where SI = 0 and BitString is of 8 bits. BPs 1, 2, 3, 4, and 5 are represented by 1 (0:00000001), 2 (0:00000010), 3 (0:00000100), 4 (0:00001000) and 5 (0:00010000) respectively.

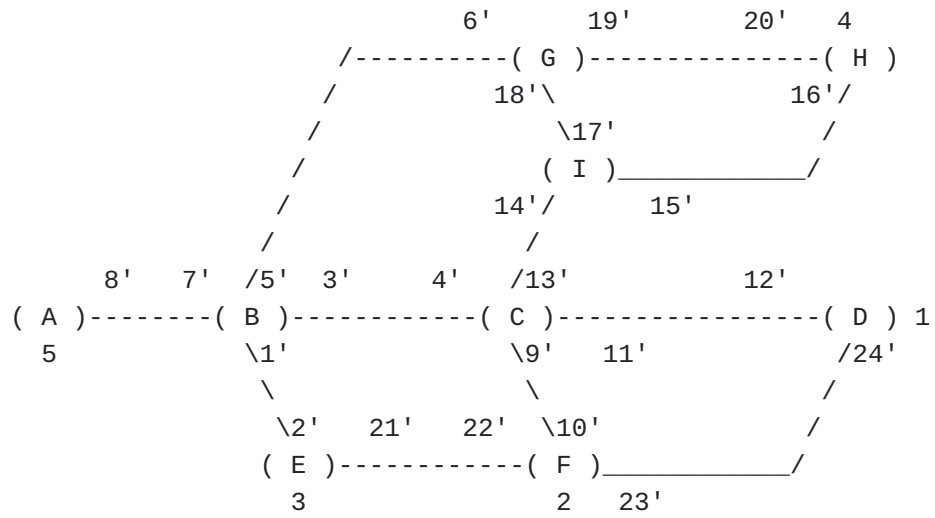

Figure 1: Example BIER-TE Topology

The BitPositions for the forward connected adjacencies are represented by i', where i is from 1 to 24. In one option, they are encoded as (n+i), where n is a power of 2 such as 32768. For simplicity, these BitPositions are represented by (SI:BitString), where SI = (6 + (i-1)/8) and BitString is of 8 bits. BitPositions i' (i from 1 to 24) are represented by 1'(6:00000001), 2'(6:00000010), 3'(6:00000100), 4'(6:00001000), 5'(6:00010000), 6'(6:00100000),

7'(6:01000000), 8'(6:10000000), 9'(7:00000001), 10'(7:00000010), . . . , 24'(8:10000000).

For a link between two nodes X and Y, there are two BitPositions for two forward connected adjacencies. These two forward connected adjacency BitPositions are assigned on nodes X and Y respectively. The BitPosition assigned on X is the forward connected adjacency of Y. The BitPosition assigned on Y is the forward connected adjacency of X.

For example, for the link between nodes B and C in the figure, two forward connected adjacency BitPositions 3' and 4' are assigned to two ends of the link. BitPosition 3' is assigned on node B to B's end of the link. It is the forward connected adjacency of node C. BitPosition 4' is assigned on node C to C's end of the link. It is the forward connected adjacency of node B.

2.2. BIER-TE Path with Multiple BitStrings

One BIER-TE path is the explicit multicast P2MP path from ingress A to egresses D and F, traversing from A to B to C, and from C to D and F. This path is represented by BPs as {7', 4', 12', 10', 1, 2}, which is {7'(6:01000000), 4'(6:00001000), 12'(7:00001000), 10'(7:00000010), 1(0:00000001), 2(0:00000010)}. These six bit positions on the path are in three sets of bit strings with SI = 0, 6 and 7.

Bit positions 1 and 2 are in the set with SI = 0, which is (0:00000011). Bit positions 7' and 4' are in the set with SI = 6, which is (6:01001000). Bit positions 12' and 10' are in the set with SI = 7, which is (7:00001010).

At ingress A, the packet to be transported by the path must be encapsulated in a BIER-TE header containing all three sets of bit strings. These sets represent the bit positions {7', 4', 12', 10', 1, 2} on the path.

The packet with the BIER-TE header is delivered from ingress A to BFR B using bit position 7' with SI = 6 in the header. BFR B forwards the packet to BFR C using bit position 4' with SI = 6 in the header. BFR C forwards a copy of the packet to BFER D using bit position 12' with SI = 7 in the header and another copy to BFER F using bit position 10' with SI = 7 in the header. BFER D decapsulates the packet and sends the payload of the packet to the packet's nextproto within BFER D using bit position 1 with SI = 0 in the header. BFER F decapsulates the packet and sends the payload of the packet to the packet's nextproto within BFER F using bit position 2 with SI = 0 in the header.

If a BIER-TE header can contain only one set of bit strings, the packet to be transported by the path cannot be delivered to the egresses of the path. At ingress A, three copies of the packet to be transported by the path are produced. Each copy contains a header with a set of bit strings. The first copy has a header with set of bit strings (0:00000011) for bit positions 1 and 2. The second copy has a header with set of bit strings (6:01001000) for bit positions 7' and 4'. The third copy has a header with set of bit strings (7:00001010) for bit positions 12' and 10'.

For the first copy, ingress A will drop it since bit positions 1 and 2 are not any adjacency bit position of A. Similarly, ingress A will the third copy since bit positions 12' and 10' are not any adjacency bit position of A.

For the second copy, ingress A sends it to BFR B using bit position 7' in the header. After receiving the packet, BFR B sends the packet to BFR C using bit position 4' in the header. After receiving the packet, BFR C drops it since there is no bit position of BFR C in the header of the packet.

3. Extensions for Multiple BitStrings

This section describes a BIER-TE header containing multiple BitStrings with different SIs (or multiple BitStrings for short), two levels of BIFTs for efficient processing the packets with the BIER-TE header, and a forwarding procedure for handling the packets using the two levels of BIFTs.

3.1. BIER-TE Header with Multiple BitStrings

A BIER-TE header needs to contain multiple sets of bit strings (i.e., multiple BitStrings with different SIs) for a BIER-TE path. In one option, they are represented by n indicating the number of BitStrings, and a pair of SI and BitString for each of the n sets of bit strings: SI-1, BitString-1; SI-2, BitString-2; ...; SI-n, BitString-n.

Figure 2 illustrates a format of a BIER-TE header having multiple BitStrings.

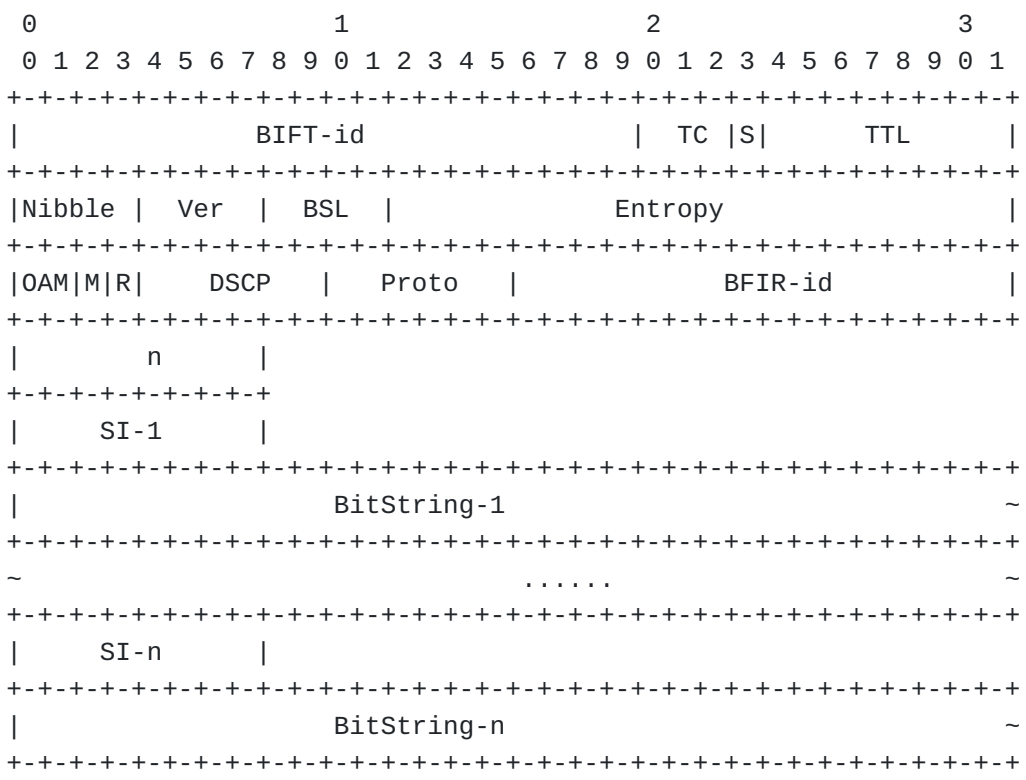

Figure 2: A Format of BIER-TE Header

BIFT-id: This field indicates a BIFT.

M: This one bit flag is set to 1 for the header containing multiple sets of bit strings, 0 for for the header not containing multiple sets of bit strings.

R: This R bit is currently unused. It SHOULD be set to 0 upon transmission and MUST be ignored upon reception.

n: It indicates the number of sets of bit strings in the header.

SI-1: It is the set identifier of the first (1-th) bit string.

BitString-1: It is the first (1-th) bit string. The length of the bit string is indicated by BSL.

SI-n: It is the set identifier of the n-th bit string.

BitString-n: It is the n-th bit string. The length of the bit string is indicated by BSL.

The other fields are the same as those in [RFC8296].

3.2. Two Levels of BIFTs

A BFR has two levels of BIFTs for BIER-TE. At the top or first level, there is one BIFT. The structure of this BIFT is shown in Figure 3. This top level BIFT has an entry for every set identifier (SI). The entry contains:

- **BitString:** The bit string (i.e., the adjacency bit positions) of the BFR in the set indicated by SI.

- **Pointer to 2nd Level BIFT:** Pointer to the second level BIFT for the bit string of the BFR in the set indicated by SI. If the bit string is all zeros, there is no second level BIFT for it and the pointer is NULL.

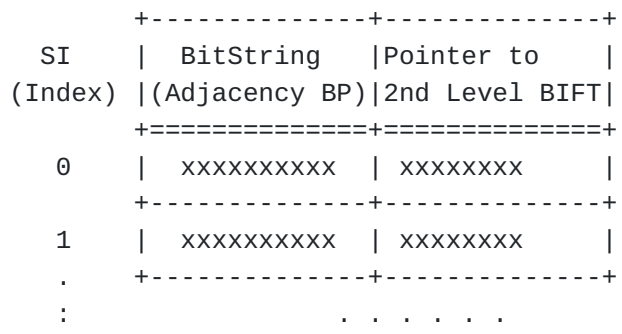

Figure 3: Structure of Top Level BIFT

For example, the top level BIFT on BFR E is illustrated in Figure 4. There are 9 sets of bit strings in total in the BIER-TE network in Figure 1. So, the BIFT has 9 entries.

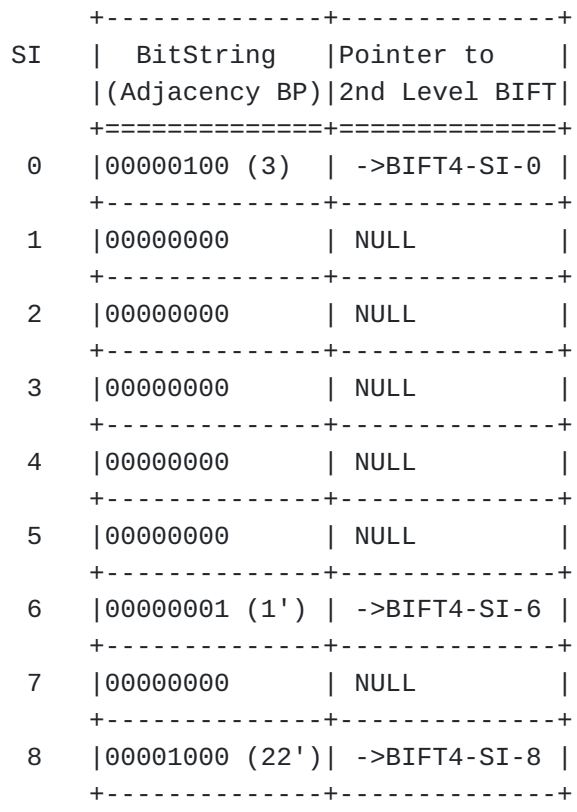

Figure 4: Top Level BIFT on BFR E

The first entry is for the set of bit string (i.e., adjacency bit positions) with SI = 0. It contains:

o BitString = 00000100. It indicates the local decap adjacency Bit Position 3 of BFR E.

o Pointer to 2nd Level BIFT = ->BIFT4-SI-0. It is a pointer to the second level BIFT for the bit string with SI = 0.

The second entry is for the set of bit string (i.e., adjacency bit positions) with SI = 1. It contains 00000000 and NULL for BitString and Pointer to 2nd Level BIFT respectively. BitString = 00000000 means that BFR E has no adjacency bit position in the set with SI = 1.

The 3-th to 6-th entries and the 8-th entry are similar to the second entry.

The 7-th entry is for the set of bit string (i.e., adjacency bit positions) with SI = 6. It contains:

o BitString =
00000001. It indicates the forward-connected adjacency Bit Position 1' of BFR E.

o Pointer to 2nd Level BIFT = ->BIFT4-SI-6. It is a pointer to the second level BIFT for the bit string with SI = 6.

The 9-th entry is for the set of bit string (i.e., adjacency bit positions) with SI = 8. It contains:

o BitString = 00001000. It indicates the forward-connected adjacency Bit Position 22' of BFR E.

o Pointer to 2nd Level BIFT = ->BIFT4-SI-8. It is a pointer to the second level BIFT for the bit string with SI = 8.

A second level BIFT for the bit string identified by a SI contains the entries for the adjacency bit positions (or say bit string) in the set identified by the SI. Its structure is shown in Figure 5. It is the same as the BIFT in [RFC9262].

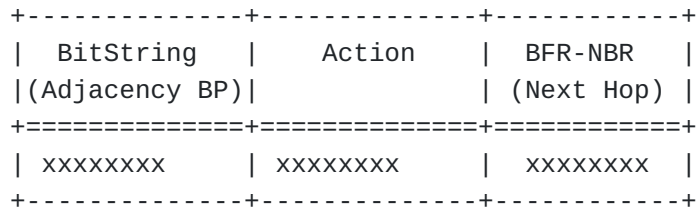

Figure 5: Structure of Second Level BIFT for SI

For example, BFR E has three adjacency bit positions: 3, 1' and 22'. They are in the three sets of bit strings identified by SI = 0, 6 and 8 respectively. So, BFR E has three second level BIFTs: BIFT for SI = 0, BIFT for SI = 6 and BIFT for SI = 8. These BIFTs are illustrated in Figure 6, Figure 7 and Figure 8.

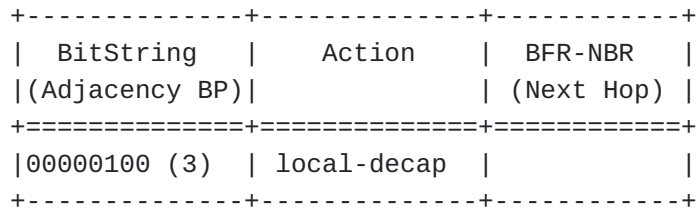

Figure 6: BIFT for SI = 0 on BFR E

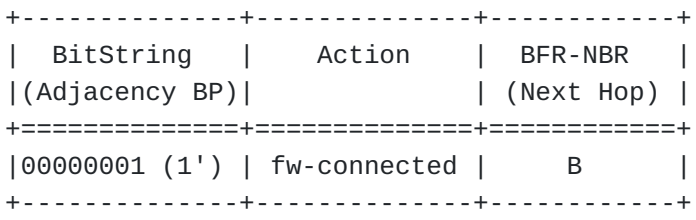

Figure 7: BIFT for SI = 6 on BFR E

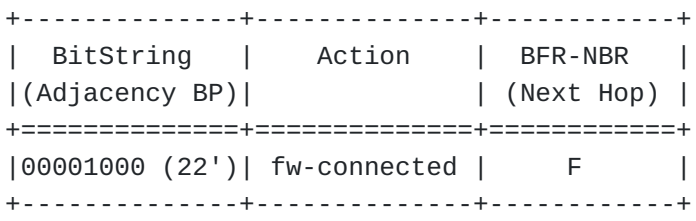

Figure 8: BIFT for SI = 8 on BFR E

In another example, BFR B has four adjacency bit positions: 2', 4', 6' and 8'. They are in the same set of bit strings identified by SI = 6. So, BFR B has one second level BIFT: BIFT for SI = 6. This BIFT is illustrated in Figure 9.

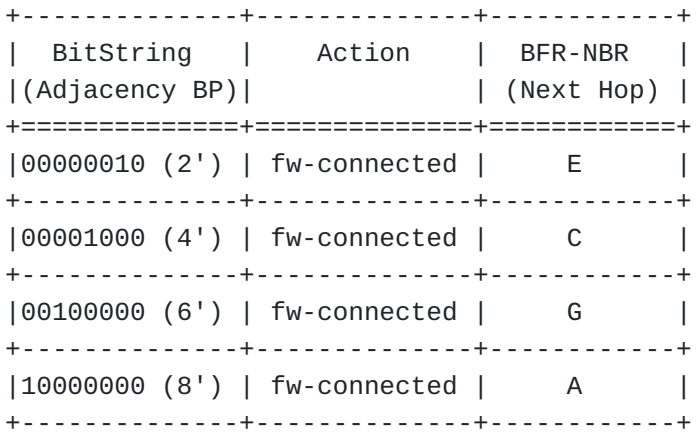

Figure 9: BIFT for SI = 6 on BFR B

The top level BIFT on BFR B is shwon in Figure 10. There are 9 sets of bit strings in total in the BIER-TE network in Figure 1. So, the BIFT has 9 entries.

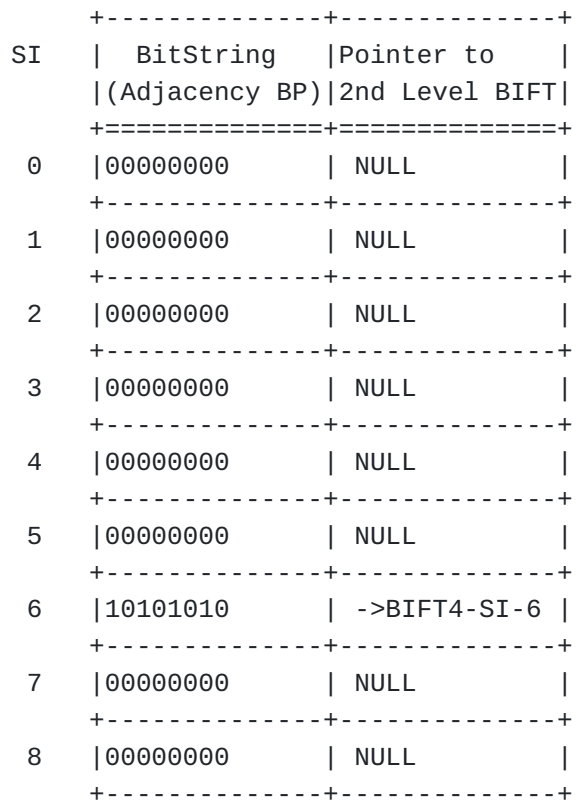

Figure 10: Top Level BIFT on BFR B

The 7-th entry is for the set of bit string (i.e., adjacency bit positions) with SI = 6. It contains:

- **BitString =** 10101010. It indicates the forward-connected adjacency Bit Position 2', 4', 6' and 8' of BFR B.

- **Pointer to 2nd Level BIFT =** ->BIFT4-SI-6. It is a pointer to the second level BIFT for the bit string with SI = 6.

The other entries are for the sets of bit strings (i.e., adjacency bit positions) with SI other than 6. Each of them contains 00000000 and NULL for BitString and Pointer to 2nd Level BIFT respectively. BitString = 00000000 means that BFR B has no adjacency bit position in the set with SI other than 6.

3.3. Forwarding Procedure

For a packet with a BIER-TE header containing multiple BitStrings with different SIs, after receiving the packet, a BFR checks each BitString to see if it has any adjacency bit positions of the BFR.

If a BitString contains an adjacency bit position of the BFR, the BFR processes the packet according to the adjacency bit position. If the adjacency bit position is a forward-connected adjacency, the BFR forwards a packet copy to the adjacency. If the adjacency bit

position is a local decap adjacency, the BFR sends the packet payload to the packet's NextProto within the BFR. This is the same as the existing behavior.

For a BitString identified by SI-i and BitString-i, the BFR determines if it contains an adjacency bit position of the BFR using the top level BIFT. The BFR gets its adjacency bit positions in the set SI-i from the BIFT and checks if BitString-i and the bit positions have the same bit with value 1. This can be achieved by checking if BIFT[SI-i][0] AND BitString-i is not zero, where BIFT[SI-i][0] is the adjacency bit positions of the BFR in the set SI-i, AND is bit wise logical and.

When BitString-i contains an adjacency bit position of the BFR, the BFR processes the packet using the second level BIFT for its adjacency bit positions in the BitString identified by SI-i. The BFR gets the second level BIFT from the top level BIFT using SI-i. The second collumn of the row with index SI-i in the top level BIFT (i.e., BIFT[SI-i][1]) stores a pointer to the second level BIFT.

For each adjacency bit position of the BFR in the BitString, the BFR processes the packet using the second level BIFT pointed by BIFT[SI-i][1] in the same way as the existing one.

The procedure for processing a BIER-TE packet is described in Pseudo code in Figure 11.

```
Packet = the packet received by BFR;
FOR i = 1 to n {// n in header is number of BitStrings
   T = BIFT[SI-i][0] & BitString-i;
   IF (T) {//has an adjacency BP of BFR
       BIFT4-SI-i = BIFT[SI-i][1]; //Get second level BIFT
       get m; //m: number of adjacency BPs in set SI-i or BIFT4-SI-i
       FOR (j = 1; T && j < m; j++) {//for each BP of BFR in set SI-i
          IF (T & BIFT4-SI-i[j][0]) {//has adjacency BP at j
             IF (BIFT4-SI-i[j][1] == fw-connected){//fw-connected adj
                 send a packet copy to BIFT4-SI-i[j][2];
             } ELSE IF (BIFT4-SI-i[j][1] == local-decap) {//decap adj
                 send packet payload to multicast overlay;
             }
             T = T & ~(BIFT4-SI-i[j][0])//Clear T's corresponding bit
          }
       }
   }
}
```

Figure 11: Forwarding Procedure for Processing BIER-TE Packet

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>.

[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>.

[RFC9262] Eckert, T., Ed., Menth, M., and G. Cauchie, "Tree Engineering for Bit Index Explicit Replication (BIER-TE)", RFC 9262, DOI 10.17487/RFC9262, October 2022, <https://www.rfc-editor.org/info/rfc9262>.

7.2. Informative References

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, DOI 10.17487/RFC3032, January 2001, <https://www.rfc-editor.org/info/rfc3032>.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, DOI 10.17487/RFC5462, February 2009, <https://www.rfc-editor.org/info/rfc5462>.

[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

Huaimo Chen
Futurewei
Boston, MA,
United States of America

Email: Huaimo.chen@futurewei.com

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.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

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

Email: wangaj3@chinatelecom.cn

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

IBM Corporation
United States of America

Email: xufeng.liu.ietf@gmail.com