

Network Working Group
Internet-Draft
Intended status: Experimental
Expires: September 6, 2018

R. Huang
T. Eckert
N. Wei
Huawei
P. Thubert
Cisco
March 5, 2018

Encapsulation for BIER-TE
draft-huang-bier-te-encapsulation-00

Abstract

This document proposes an enhanced encapsulation for BIER to support BIER, BIER-TE and a control word.

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 September 6, 2018.

Copyright Notice

Copyright (c) 2018 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	2
2.	BIER-TE Encapsulation (normative)	3
2.1.	BT bit - Simultaneous support for BIER and BIER-TE	3
2.2.	BIFT-ID	3
2.3.	Control Word and flows	3
2.4.	Header format & fields	4
3.	BIER-TE based resilience operations (informational)	5
4.	BitStringLength (BSL) considerations (informational)	6
4.1.	IPTV	7
4.2.	Multicast in L3VPN	8
5.	Acknowledgements	9
6.	Security Considerations	9
7.	IANA Considerations	9
8.	References	9
8.1.	Normative References	9
8.2.	Informative References	10
	Authors' Addresses	10

[1.](#) Introduction

[BIER-TE-ARCH] specifies BIER-TE: Traffic Engineering for Bit Index Explicit Replication (BIER). It builds on the BIER architecture as described in [RFC8279](#) [[RFC8279](#)], but uses every BitPosition of the BitString of a BIER-TE packet to indicate one or more adjacencies instead of a BFER as in BIER.

This document proposes an enhanced version of the MPLS and non-MPLS encapsulation for BIER packets to support both BIER and BIER-TE. It is based on [RFC8296](#) [[RFC8296](#)].

This enhanced encapsulation also adds support for a control word in the header and discusses it. Finally, the document discusses BitStringLength (BSL) size requirements in implementations for informational reasons to help aid implementors to determine an appropriate BSL.

[1.1.](#) Terminology

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](#) [[RFC2119](#)].

2. BIER-TE Encapsulation (normative)

2.1. BT bit - Simultaneous support for BIER and BIER-TE

This document supports mixed BIER and BIER-TE forwarding in a domain. Either or both of them may be used in a domain. The overall solution to support this depends on additional signaling such as existing BIER ISIS/BGP signaling. Architecturally, every SD SHOULD only use a single Type of BIER: BIER or BIER-TE. Note that this document will use the abbreviation BT to refer to the Bier Type.

In the presence of BIER and BIER-TE together in the network, there is always a risk of receiving a packet which is meant to be of one BT and processing it through a BIFT of the other BT. This can come from misconfiguration even in the face of signalling via IGP/BGP. The risk increases also when packets are generated modular from applications on PE or other sources and could use both BTs. To resolve this, the header includes a bit to indicate the BT. If the BT of a packet is inconsistent with the BT of the BIFT on the BFR, the BFR MUST NOT forward it. OAM actions MAY be triggered (subject to future work).

Note that the TTL field of the existing BIER packet header (or of IP packets) spends 7 bits on loop prevention. One bit for the BT is a comparably low cost to protect against a similar degree of problems.

Indicating the BT explicitly through a bit in the encapsulation is called the "explicit" option. Relying solely on the BT of the BIFTs is called "implicit" option. In this version of the document, we choose the explicit option for reasons outlined above.

2.2. BIFT-ID

Like in the original BIER header, the semantic of the BIFT-id of the header is that it is representing a <SI,SD,BSL> on the BFR receiving the packet. In the case of MPLS forwarding, the expectation is that the protocols to signal label ranges would be extended to also signal label ranges for the SD using BIER-TE. This is subject to the work of other documents. In the case of non-MPLS forwarding, no additional signaling may be necessary, and BIER and BIER-TE packets using the encapsulation of this document would equally use the BIFT-ID encoding as described in [[BIER-non-MPLS](#)].

2.3. Control Word and flows

This document adds a "control word" to the BIER packet header to allow that BIER or BIER-TE packets with this header could be used as

a DetNet Data Plane, independent of MPLS encapsulation, see [\[I-D.ietf-detnet-dp-sol\]](#), section 5.3 (in revision 01).

The control word provides a sequence number, therefore allowing to correct reordering and discover packet loss. The primary use though is resilient dual-path transmission of two copies of the same packet via disjoint paths. This is specifically a desirable use-case with BIER-TE because it allows the engineering of such disjoint paths. The flow to which the sequence number in the control word applies is <BFR-id,entropy>.

Note: The justification to carry a control word in the BIER encapsulation is similar to carrying the BFIR-ID in it. Initially, both could be seen as primarily required on BIER domain edge-nodes as part of the overlay using BIER, but not by BIER/BIER-TE itself directly. See [Section 3](#) for more explanation how the resilience mechanism requiring the control word would work. Compared to the BFIR-ID, there is also the option to leverage it within BIER-TE itself. The details of that operations is subject to other specifications.

The authors think that the overhead of the control word is always acceptable for BIER-TE. For BIER, the use of this extended header version is optional, therefore BIER packets that need a control word would use this version of the header, those that do not need it would use version 1. If this overhead is considered to not be acceptable for all BIER-TE packets, the encoding could make those 32 bits optional through the use of one of the reserved bits or version numbers or by using a bit in the header to indicate whether the control word is present or not.

[2.4.](#) Header format & fields


```

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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                               BIFT-id                               | TC |S|   TTL   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|Nibble| Ver | BSL |                               Entropy (flow-id) |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|OAM|T|R|   DSCP   |   Proto   |                               BFIR-id   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0 0 0 0|   Control Word (sequence number)                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                               BitString (first 32 bits)                               ~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
~                               BitString (last 32 bits)                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
0                               1                               2                               3

```

All header fields not described below are left unchanged from [\[BIER-TE-ARCH\]](#)

T: This 1-bit field identifies that the packet is to be forwarded as a BIER packet (0) or a BIER-TE packet(1).

Ver: The version of this header format is 2.

R: Reserved - unchanged (just reduced by one bit from version). Must be set to 0.

Entropy: Unchanged, but double-used as part of the flow-identifier together with the control word

Control Word: The control word in the terminology of MPLS pseudowires (where it originates from) is the full 32 bits. For detnet, the current target is 28 bits of sequence number and 4 bits 0 preceeding it.

3. BIER-TE based resilience operations (informational)

This section discusses how resilience operations with the help of the sequence number in the control word of the header in this document can be operated as an overlay (BFIR-BFER) function but also points out that it could become an integral (optional) part of BIER-TE itself. This section is solely informational. The planned document

to describe the BIER-TE forwarding aspects of resilience operations is [[I-D.thubert-bier-replication-elimination](#)].

The BFIR determines - potentially with the help of a BIER-TE Controller Host (controller) - a bitstring that forms two disjoint DAGs (Directed Acyclic Graphs) through the BIER-TE domain towards the same set of BFER. In addition, an entropy value is decided (by BFIR and/or controller) and signalled to the BFER. The BFER can therefore set up "duplicate elimination state": The BFIR increments the sequence number with every packet of the flow it sends. The BFER assign packets to a flow by <bfir-id,entropy> and perform duplicate elimination on them.

Note that the bitstring as seen on the receiving BFER can provide additional diagnostics, for example the bits not reset by forwarding in BIER-TE give an indication about which path the BIER-TE packet was forwarded.

Instead of simply considering this protected mode of operations solely an end-to-end (BFIR/BFER) function, it could also be more flexibly embedded into BIER-TE itself, allow to provide in-BIER-TE segmented Packet Replication and (duplicate) Elimination Functions (PREF) definable by the bitstring of a BIER-TE packet. This could be achieved by adding to BIER-TE forwarding functions new adjacency types for duplication with sequence-number generation and duplicate-elimination. The ability to perform such processing as part of BIER-TE itself is the primary reason to ensure that all the necessary elements for such operations are part of the BIER-TE header itself.

4. BitStringLength (BSL) considerations (informational)

BIER-TE uses each BitPostion to indicate the adjacencies instead of a BFER as in BIER, it therefore consumes more BitPostions than BIER. In BIER-TE, the number of adjacencies passed by one BIER-TE packet MUST be less than the value of BitString length (BSL). The BIER-TE architecture discusses a range of options to reduce the number of bits for intermediate hops through various BIER-TE adjacencies and how to use them.

The maximum supported BSL has a different impact in BIER-TE than it has in BIER: A smaller maximum supported BSL in BIER primarily leads to less replication efficiency: With a BSL of 256, BIER can be up to 256 more efficient than unicast (1 packet for 256 receivers). In BIER-TE, the BSL also limits the size of the topology towards BFER and the alternative paths that can explicitly be engineered to reach the BFER. One simple guess is that 50% of bits in a bitstring may be required for intermediate hops, therefore requiring about double the amount of bits as BIER - as the cost of being able to engineer pats.

So far, there is no comprehensive analysis of the number of required bits for specific scenarios in BIER-TE. The following subsections give two examples of such scenarios and how to use and save BIER-TE bits for intermedia hops.

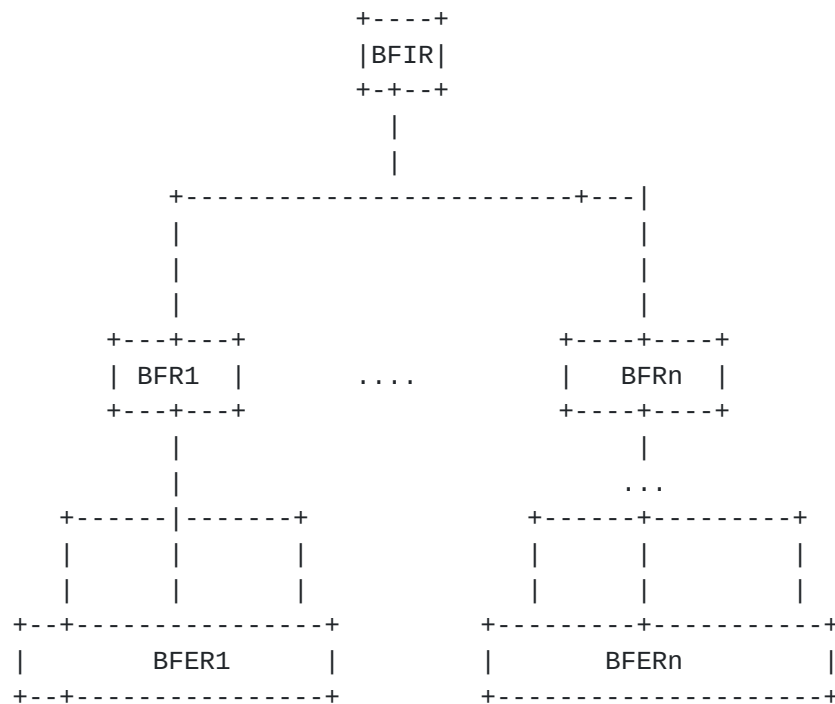
4.1. IPTV

Multicast is widely used for IPTV services by simultaneously delivering a single stream of video to thousands of recipients. Currently, PIM is widely used to provide multicast capability usually from core router(CR) to provider edges (PEs). And the multicast tree is usually constructed in the hierarchical way. The end users using PIM/IGMP to request the multicast data. BIER can be well used in from CR to those PEs. The number of hops from multicast source (CR) which could be BFIRs, to the multicast receivers (PE) which can be regarded as BRERs is usually no more than 10. BIER-TE will be useful in the cases where different video channels can have different transport paths to achieve load balancing.

To save the bit consumption, 2 ways could be used:

1. Multiple BFRs and routes are required to receive the same data. These BFRs or links can share one bit.
2. Different bits can be used for pruning. But these bits can be reused in similar but different groups.

Considering an example illustrated as following:



BFR1 and BFRn, and other BFRs in between, can share one bit because they are receiving all the content and don't need pruning. From BFR1 to BFER1, there are 3 ways to reach BFER1. So these 3 ways can be assigned with different bits. But these 3 bits can be reused in the group from BFRn to BFERn, and other groups in between, which share the same topology as the group from BFR1 to BFER1.

BIER-TE can be well implemented using these 2 ways to save the bit consumption in IPTV networks with the similar topologies like the above example.

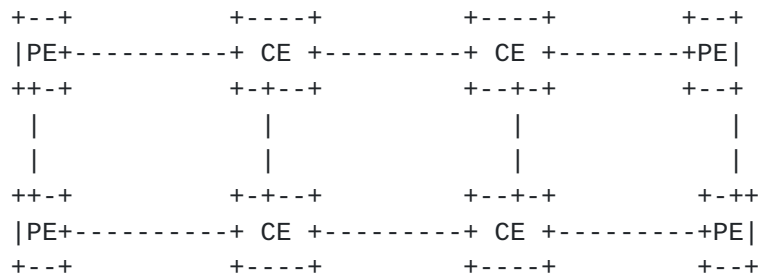
4.2. Multicast in L3VPN

MVPN is a technology to deploy multicast service in an existing VPN or as part of a transport infrastructure. Multicast data is transmitted between private networks over a VPN infrastructure by encapsulating the original multicast packets. PE routers are connected to these private networks either containing receivers or senders.

There are several multicast applications widely using the MVPN deployment. For example, L3VPN multicast service offered by service providers to enterprise customers, and video transport applications for separation between different customers: One content provider may provide video wholesale service to another, or multiple content providers may share one network to transport video from headend. Especially the latter case, network SLAs should be guaranteed as the

original video content is precious. Thus, traffic engineering is required.

According to the current implementations, the scale of a MVPN network usually contains less than several hundreds of PEs, and hundreds of core routers which are connected in full mesh, like following figure illustrated.



In such a case, the ways in Section 7.5.2 of [\[BIER-TE-ARCH\]](#) can be used by regarding the CE area as the Core. Based on this, current BIER design is sufficient to be reused in BIER-TE.

5. Acknowledgements

TBD.

6. Security Considerations

The security considerations are in compliance with BIER-TE architecture [\[BIER-TE-ARCH\]](#) and BIER encapsulation [RFC8296](#) [\[RFC8296\]](#). And the content in this document does not create any other attacks or security concerns.

7. IANA Considerations

TBD.

8. References

8.1. Normative References

[\[BIER-non-MPLS\]](#)

Wijnands, I., Xu, X., and H. Bidgoli, "An Optional Encoding of the BIFT-id Field in the non-MPLS BIER Encapsulation", ID [draft-wijnandsxu-bier-non-mpls-bift-encoding-01](#) (work in progress), August 2017.

[BIER-TE-ARCH]

Eckert, T., Cauchie, G., Braun, W., and M. Menth, "Traffic Engineering for Bit Index Explicit Replication BIER-TE", ID [draft-ietf-bier-te-arch-00](#) (work in progress), January 2018.

[I-D.thubert-bier-replication-elimination]

Thubert, P., Eckert, T., Brodard, Z., and H. Jiang, "BIER-TE extensions for Packet Replication and Elimination Function (PREF) and OAM", [draft-thubert-bier-replication-elimination-03](#) (work in progress), March 2018.

[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>>.

[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>>.

8.2. Informative References**[I-D.ietf-detnet-dp-sol]**

Korhonen, J., Andersson, L., Jiang, Y., Finn, N., Varga, B., Farkas, J., Bernardos, C., Mizrahi, T., and L. Berger, "DetNet Data Plane Encapsulation", [draft-ietf-detnet-dp-sol-01](#) (work in progress), January 2018.

Authors' Addresses

Rachel Huang
Huawei
101 Software Avenue, Yuhua District
Nanjing 210012
China

Email: rachel.huang@huawei.com

Toerless Eckert
Huawei USA - Futurewei Technologies Inc.
2330 Central Expy
Santa Clara 95050
USA

Email: tte+ietf@cs.fau.de

Naiwen Wei
Huawei

Email: weinaiwen@huawei.com

Pascal Thubert
Cisco Systems
Village d'Entreprises Green Side
400, Avenue de Roumanille
Batiment T3
Biot - Sophia Antipolis 06410
FRANCE

Phone: +33 4 97 23 26 34
Email: pthubert@cisco.com

