BIER Split-horizon mechanism for active-active access
draft-hao-bier-active-active-00.txt

Abstract

This document proposes a BIER split-horizon mechanism for active-active access. Both data plane and control plane extension are included.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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.”

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Copyright Notice

Copyright (c) 2015 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 (http://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

Hao & Li, etc Expires March 9, 2016 [Page 1]
Bit Index Explicit Replication (BIER) ([BIER-ARCH]) is an architecture that provides optimal multicast forwarding through a "multicast domain", without requiring intermediate routers to maintain any per-flow state. BIER doesn’t engage in any explicit tree-building protocols, it can be thought of implicitly created P2MP tunnels from each "Bit Forwarding Ingress Router" (BFIR) to all the "Bit Forwarding Egress Routers" (BFERs) in the BIER domain. BIER is potentially applicable to many services where Multicast is used, especially as P-Tunnel for MVPN, EVPN, etc.

```
+------+
|  CE4 |
+------+
      |
+--------+
|   BFR4 |
+--------+

      |
------ |  |
------ |  |
------ |  |
      +---+  +---+  +---+
      |  |    |  |    |  |
      |  BFR1 |  BFR2 |  BFR3 |
      +---+  +---+  +---+
      *    |  *    |  *    |
      *    |  *    |  *    |
```

1. Introduction

Bit Index Explicit Replication (BIER) ([BIER-ARCH]) is an architecture that provides optimal multicast forwarding through a "multicast domain", without requiring intermediate routers to maintain any per-flow state. BIER doesn’t engage in any explicit tree-building protocols, it can be thought of implicitly created P2MP tunnels from each "Bit Forwarding Ingress Router" (BFIR) to all the "Bit Forwarding Egress Routers" (BFERs) in the BIER domain. BIER is potentially applicable to many services where Multicast is used, especially as P-Tunnel for MVPN, EVPN, etc.
As illustrated in figure 1, to provide redundant connectivity, a CE is normally connected to multiple BFRs through an Ethernet Link Aggregation Group with LACP [802.1AX], the CE can be a multicast source or receiver. The BFRs offering multi-homing is called Active-Active BFR (AABFR) group. In Active-Active case, the following two problems for multicast packet forwarding should be solved:

1. Duplicated delivery of flooding traffic

The issue occurs on the multi-homed BFERs connecting to one or more multicast receivers. If more than one BFER out of the AABFR group egress a copy of a multicast packet from BIER domain, the Multicast receiver will see packet duplication. Therefore, it is REQUIRED that a unique BFER is appointed as the multicast egress BFR. This unique multicast egress NVE can be appointed according to a unanimous agreed designated forwarder (DF) election algorithm by all BRRs in a AABFR group. The election algorithm should be used to ensure service carving balance among different BFRs in a redundant group, the service carving factor can be based on VPN instance, multicast group, access VLAN, etc, so it is carried service dependant rather than service agnostic. The DF election process defined in [EVPN] is an example for DF PE determination.

2. Loop & Echo Forwarding among multi-homed PEs.

The issue occurs on the multi-homed BFIRs connecting to one or more multicast sources. Assuming a multicast source S1 is multi-homed to BFIR1 and BFIR2 using an Ethernet Link Aggregation Group with LACP [802.1AX], when S1 sends a multicast packet to BFIR1, the BFIR1 will inject the packet to BIER domain and the packet will be received by BFIR2 from the BIER domain, then BFIR2 will perform BIER decapsulation and send the packet to S1 again. Split-horizon filtering mechanism should be used to prevent the loop between BFIR1 and BFIR2. The split-horizon mechanism relies on the packet from ingress device carrying origin identification to identify the segment from which the frame entered the BIER network, the egress
device performs filtering function based on the origin identification. The split-horizon mechanism should be service agnostic and should be maintained in transport layer, i.e, BIER should provide the Split-horizon capability.

In summary, BIER as a transport layer technology, it should provide split-horizon capability for active-active access and can be shared by all multicast services. As for DF election mechanism, it should be provided in each multicast service. This document will define the data plane and control plane extension respectively for BIER split-horizon mechanism.

2. Conventions used in this document

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].

The terms and acronyms in [RFC6325] are used with the following additions:

- Ethernet Segment (ES): When a customer site (device or network) is connected to one or more PEs via a set of Ethernet links, then that set of links is referred to as an 'Ethernet segment'.

- Ethernet Segment Identifier (ESI): A unique non-zero identifier that identifies an Ethernet segment is called an 'Ethernet Segment Identifier'.

- MVPN: Multicast Virtual Private Network -- a VPN [RFC4364] in which multicast service is offered.

- P-tunnel. A multicast tunnel through the network of one or more SPs. P-tunnels are used to transport MVPN multicast data.

3. Solution overview

Every BFR track the BFID(es) associated with the other BFR(s) with which it has shared multi-homed Ethernet Segments. When the BFIR receives a multi-destination frame from the BEIR domain, it examines the source BFID in the BIER header (which corresponds to the BFIR) and filters out the frame on all local interfaces connected to Ethernet Segments that are shared with the ingress BFR.

Each BFIR performs replication locally to all directly attached Ethernet Segments for all flooded traffic ingress from the access
interfaces (i.e. from the multicast source). This approach is referred to as "Local Bias".

4. Data plane format

In [BIER-MPLS], the BIER header format is defined as follows:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0 1 0 1|  Ver |  Len |              Entropy                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                BitString  (first 32 bits)                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                BitString  (last 32 bits)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|OAM|     Reserved      | Proto |            BFIR-id            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: BIER Header

BFIR-id field is the BFR-id of the BFIR, it can be used for split-horizon filtering.

5. Discovery of Multi-homing BFRs

Each Ethernet segment is assigned a global ESI as Ethernet segment identification. The ESI can be manually configured or automatically derived.

A couple of different mechanisms are available to derive the ESI automatically, such as snooping LACP packets or LLDP packets. Once the ESI for an Ethernet segment is assigned for a multi-homed CE, it is advertised by the BFRs through IS-IS or OSPF extension.

The following two ESI values are reserved:

- ESI 0 denotes a single-homed site.

- ESI {0xFF} (repeated 10 times) is known as MAX-ESI and is reserved.

In general, an Ethernet segment SHOULD have a non-reserved ESI that is unique network wide (i.e., across all EVPN instances on all the PEs).
5.1. IS-IS extension

BIER Info sub-TLV (Section 6.1 [BIER-ISIS]) carries the information for the BIER sub-domains that the router participates in as BFR. A newly defined BIER ESI sub-sub-TLV carries the information of local ESIs and is carried within the BIER Info sub-TLV that the router participates in as BFR. The BIER ESI sub-sub-TLV format is as follows:

```
+-----------------+
| Type(1 octet)   |
| Length(1 octet) |
+-----------------+
```

```
+------------------------------+
| Ethernet Segment Identifier  |
| (1) (10 octets)             |
+------------------------------+
```

```
| ... ...                      |
+------------------------------+
```

```
+------------------------------+
| Ethernet Segment Identifier  |
| (N) (10 octets)             |
+------------------------------+
```

Type: TBD
Length: 1 byte

Ethernet Segment Identifier: It is encoded as a 10-octet integer in line format with the most significant octet sent first.

5.2. OSPF extension

BIER ESI Sub-TLV is a sub-TLV of the BIER Sub-TLV [BIER-OSPF]. BIER ESI Sub-TLV is used in order to advertise local connecting ESI used for BIER. The BIER ESI sub-TLV format is as follows:

```
+-----------------+
| Type(2 octet)   |
| Length(2 octet) |
+-----------------+
```

```
+------------------------------+
| Ethernet Segment Identifier  |
| (1) (10 octets)             |
+------------------------------+
```

```
| ... ...                      |
+------------------------------+
```

```
+------------------------------+
| Ethernet Segment Identifier  |
| (N) (10 octets)             |
+------------------------------+
```

Type: TBD
Length: 1 byte

Ethernet Segment Identifier: It is encoded as a 10-octet integer in line format with the most significant octet sent first.
6. Security Considerations

Implementations must assure that malformed TLV and Sub-TLV permutations do not result in errors which cause hard protocol failures.

7. Normative References


Acknowledgments

The authors wish to acknowledge the important contributions of Eric Wu.

Authors’ Addresses

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012, China
Email: haoweiguo@huawei.com