BIER   Z. Zhang
Internet-Draft   A. Przygienda
Intended status: Standards Track   Juniper Networks
Expires: December 10, 2017   A. Sajassi
                                                         Cisco Systems
                                                            J. Rabadan
                                                                 Nokia
                                                          June 8, 2017

EVPN BUM Using BIER
draft-zzhang-bier-evpn-00

Abstract

   This document specifies protocols and procedures for forwarding
   broadcast, unknown unicast and multicast (BUM) traffic of Ethernet
   VPNs (EVPN) using Bit Index Explicit Replication (BIER).

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.

Table of Contents

1. Introduction

   [RFC7432] and [[I-D.ietf-bess-evpn-overlay](#)] specify the protocols and
   procedures for Ethernet VPNs (EVPNs). For broadcast, unknown unicast
   and multicast (BUM) traffic, provider/underlay tunnels (referred to
   as P-tunnels) are used to carry the BUM traffic. Several kinds of
   tunnel technologies can be used, as specified in [[RFC7432](#)].

   Bit Index Explicit Replication (BIER) ([[I-D.ietf-bier-architecture](#)])
   is an architecture that provides optimal multicast forwarding through
   a "multicast domain", without requiring intermediate routers to
   maintain any per-flow state or to engage in an explicit tree-building
   protocol. The purpose of this document is to specify the protocols
   and procedures to transport EVPN BUM traffic using BIER.

The EVPN BUM procedures specified in [RFC7432] and extended in
[I-D.ietf-bess-evpn-bum-procedure-updates],
[I-D.ietf-bess-evpn-igmp-mld-proxy], and
[I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements] are much aligned with
MVPN procedures. As such, this document is also very much aligned
with [I-D.ietf-bier-mvpn]. For terseness, some background, terms and
concepts are not repeated here. Additionally, some text is borrowed
verbatim from [I-D.ietf-bier-mvpn].

1.1. Terminologies

o BFR: Bit-Forwarding Router.

o BFIR: Bit-Forwarding Ingress Router.

o BFER: Bit-Forwarding Egress Router.

o BFR-Prefix: An IP address that uniquely identifies a BFR and is
   routeable in a BIER domain.

o C-S: A multicast source address, identifying a multicast source
   located at a VPN customer site.

o C-G: A multicast group address used by a VPN customer.

o C-flow: A customer multicast flow. Each C-flow is identified by
   the ordered pair (source address, group address), where each
   address is in the customer's address space. The identifier of a
   particular C-flow is usually written as (C-S,C-G). Sets of
   C-flows can be identified by the use of the "C-*" wildcard (see
   [RFC6625]), e.g., (C-*,C-G).

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

o IMET Route: Inclusive Multicast Ethernet Tag Auto-Discovery route.
   Carried in BGP Update messages, these routes are used to advertise
   the "default" P-tunnel for a particular broadcast domain.

o SMET Route: Selective Multicast Ethernet Tag Auto-Discovery route.
   Carried in BGP Update messages, these routes are used to advertise
   the C-flows that the advertising PE is interested in.

o S-PMSI A-D route: Selective Provider Multicast Service Interface
   Auto-Discovery route. Carried in BGP Update messages, these
   routes are used to advertise the fact that particular C-flows are
   bound to (i.e., are traveling through) particular P-tunnels.

o PMSI Tunnel attribute (PTA). This BGP attribute carried is used
   to identify a particular P-tunnel. When C-flows of multiple VPNs
   are carried in a single P-tunnel, this attribute also carries the
   information needed to multiplex and demultiplex the C-flows.

2. Use of the PMSI Tunnel Attribute

[RFC7432] specifies that Inclusive Multicast Ethernet Tag (IMET)
routes carry a PMSI Tunnel Attribute (PTA) to identify the particular
P-tunnel to which one or more BUM flows are being assigned, the same
as specified in [RFC6514] for MVPN. [I-D.ietf-bier-mvpn] specifies
the encoding of PTA for use of BIER with MVPN. Much of that
specification is reused for use of BIER with EVPN and much of the
text below is borrowed verbatim from [I-D.ietf-bier-mvpn].

The PMSI Tunnel Attribute (PTA) contains the following fields:

o "Tunnel Type". The same codepoint that [I-D.ietf-bier-mvpn]
   requests IANA to assign for the new tunnel type "BIER" is used for
   EVPN as well.

o "Tunnel Identifier". When the "tunnel type" field is "BIER", this
   field contains two subfields. The text below is exactly as in
   [I-D.ietf-bier-mvpn]

   1 The first subfield is a single octet, containing the sub-
      domain-id of the sub-domain to which the BFIR will assign the
      packets that it transmits on the PMSI identified by the NLRI of
      the IMET, S-PMSI A-D, or per-region I-PMSI A-D route that
      contains this PTA. (How that sub-domain is chosen is outside
      the scope of this document.)

   2 The second subfield is the BFR-Prefix (see
      [I-D.ietf-bier-architecture]) of the originator of the route
      that is carrying this PTA. This will either be a /32 IPv4
      address or a /128 IPv6 address. Whether the address is IPv4 or
      IPv6 can be inferred from the total length of the PMSI Tunnel
      attribute.

o "MPLS label". For EVPN-MPLS [RFC7432], this field contains an
   upstream-assigned MPLS label. It is assigned by the BFIR.
   Constraints on the way in which the originating router selects
   this label are discussed in Section 2.2. For EVPN-VXLAN/NVGRE
   [I-D.ietf-bess-evpn-overlay], this field is a 24-bit VNI/VSID of
   global significance.

o "Flags". When the tunnel type is BIER, two of the flags in the
   PTA Flags field are meaningful. Details about the use of these
   flags can be found in [Section 2.1](#).

   * "Leaf Info Required per Flow (LIR-pF)"
      [I-D.ietf-bess-mvpn-expl-track]

   * "Leaf Info Required Bit (LIR)"

Note that if a PTA specifying "BIER" is attached to an IMET, S-PMSI
A-D, or per-region I-PMSI A-D route, the route MUST NOT be
distributed beyond the boundaries of a BIER domain. That is, any
routers that receive the route must be in the same BIER domain as the
originator of the route. If the originator is in more than one BIER
domain, the route must be distributed only within the BIER domain in
which the BFR-Prefix in the PTA uniquely identifies the originator.
As with all MVPN routes, distribution of these routes is controlled
by the provisioning of Route Targets.

2.1. Explicit Tracking

When using BIER to transport an EVPN BUM data packet through a BIER
domain, an ingress PE functions as a BFIR (see
[I-D.ietf-bier-architecture]). The BFIR must determine the set of
BFERs to which the packet needs to be delivered. This can be done in
either of two ways in the following two sections.

2.1.1. Using IMET/SMET routes

Both IMET and SMET (Selective Multicast Ethernet Tag
[I-D.ietf-bess-evpn-igmp-mld-proxy]) routes provide explicit tracking
functionality.

For an inclusive PMSI, the set of BFERs to deliver traffic to
includes the originators of all IMET routes for a broadcast domain.
For a selective PMSI, the set of BFERs to deliver traffic to includes
the originators of corresponding SMET routes.

The SMET routes do not carry a PTA. When an ingress PE sends traffic
on a selective tunnel using BIER, it uses the upstream assigned label
that is advertised in its IMET route.

Only when selectively forwarding is for all flows without tunnel
segmentation, SMET routes are used without S-PMSI A-D routes.
Otherwise, the procedures in the following section apply.

2.1.2. Using S-PMSI/Leaf A-D Routes

   There are two cases where S-PMSI/Leaf A-D routes are used as
   discussed in the following two sections.

2.1.2.1. Selective Forwarding Only for Some Flows

   With the SMET procedure, a PE advertises an SMET route for each
   (C-S,C-G) or (C-*,C-G) state that it learns on its ACs, and each SMET
   route is tracked by every PE in the same broadcast domain. It may be
   desired that SMET routes are not used to reduce the burden of
   explicit tracking.

   In this case, most multicast traffic will follow the I-PMSI
   (advertised via IMET route) and only some flows follow S-PMSIs. To
   achieve that, S-PMSI/Leaf A-D routes can be used, as specified in
   [I-D.ietf-bess-evpn-bum-procedure-updates]. The LIR bit may be set
   in the S-PMSI A-D routes, and the PEs that need to receive
   corresponding traffic will respond with a Leaf A-D route. The
   ingress PE identifies the set of BFERs to deliver traffic to
   according to the set of corresponding Leaf A-D routes received.

   The S-PMSI A-D route carries the same PTA as in the IMET route,
   except that similar to MVPN, the LIR-pF flag may be set for an
   ingress PE to request individual (C-S,C-G) or (C-*,C-G) Leaf A-D
   routes.

2.1.2.2. Tunnel Segmentation

   Another case where S-PMSI/Leaf A-D routes are necessary is tunnel
   segmentation, which is also specified in
   [I-D.ietf-bess-evpn-bum-procedure-updates], and further clarified in
   [I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements] for segmentation with
   SMET routes. This is only applicable to EVPN-MPLS.

   Similar to MVPN, the LIR-pF flag cannot be used with segmentation,
   and the S-PMSI A-D routes' PTA MUST carry an upstream assigned label
   to allow tunnel segmentation points to do label switching. The
   S-PMSI A-D routes could be proactively (re-)advertised by the ingress
   PEs or segmentation points, or could be triggered by the unsolicited
   Leaf A-D routes received from downstream.

2.2. MPLS Label in PTA

   Similar to the MVPN case in [I-D.ietf-bier-mvpn], the label
   allocation for the upstream assigned label in the PTA MUST follow the
   following rules (text borrowed verbatim from [I-D.ietf-bier-mvpn]).

Suppose an ingress PE originates two x-PMSI A-D routes, where we use
the term "x-PMSI" to mean "I-PMSI or S-PMSI". Suppose both routes
carry a PTA, and the PTA of each route specifies"BIER".

o If the two routes do not carry the same set of Route Targets
   (RTs), then their respective PTAs MUST contain different MPLS
   label values.

o If segmented P-tunnels are being used, then the respective PTAs of
   the two routes MUST contain different MPLS label values, as long
   as the NLRIs are not identical. In this case, the MPLS label can
   be used by the BFER to identify the particular C-flow to which a
   data packet belongs, and this greatly simplifies the process of
   forwarding a received packet to its next P-tunnel segment. This
   is explained further below.

When segmented P-tunnels are being used, an ABR or ASBR may receive,
from a BIER domain, an x-PMSI A-D route whose PTA specifies "BIER".
This means that BIER is being used for one segment of a segmented
P-tunnel. The ABR/ASBR may in turn need to originate an x-PMSI A-D
route whose PTA identifies the next segment of the P-tunnel. The
next segment may also be "BIER". Suppose an ABR/ASBR receives x-PMSI
A-D routes R1 and R2, and as a result originates x-PMSI A-D routes R3
and R4 respectively, where the PTAs of each of the four routes
specify BIER. Then the PTAs of R3 and R4 MUST NOT specify the same
MPLS label.

The ABR/ASBR MUST then program its dataplane such that a packet
arriving with the upstream-assigned label specified in route R1 is
transmitted with the upstream-assigned label specified in route R3,
and a packet arriving with the upstream-assigned label specified in
route R2 is transmitted with the label specified in route R4. Of
course, the data plane must also be programmed to encapsulate the
transmitted packets with an appropriate BIER header, whose BitString
is determined by the multicast flow overlay.

3. Multihoming Split Horizon

For EVPN-MPLS, [RFC7432] specifies the use of ESI labels to identify
the ES from which a BUM packet originates. A PE receiving that
packet from the core side will not forward it to the same ES. The
procedure works for both Ingress Replication (IR) and RSVP-TE/mLDP
P2MP tunnels, using downstream- and upstream-assigned ESI labels
respectively. For EVPN-VXLAN/NVGRE, [I-D.ietf-bess-evpn-overlay]
specifies local-bias procedures, where a PE receiving a BUM packet
from the core side knows from encapsulation the ingress PE so it does
not forward the packet to any multihoming ESes that the ingress PE is

on, because the ingress PE already forwarded the packet to those
ESes, regardless of whether the ingress PE is a DF for those ESes.

With BIER, the local-bias procedure still applies for EVPN-VXLAN/
NVGRE as the BFIR-id in the BIER header identifies the ingress PE.
For EVPN-MPLS, ESI label procedures also still apply though two
upstream assigned labels will be used (one for identifying the
broadcast domain and one for identifying the ES) - the same as in the
case of using a single P2MP tunnel for multiple broadcast domains.
The BFIR-id in the BIER header identifies the ingress PE that
assigned those two labels.

Details for split-horizon in case of segmentation will be provided in
future revisions.

4. Data Plane

Similar to MVPN, the EVPN application plays the role of the
"multicast flow overlay" as described in
[I-D.ietf-bier-architecture].

4.1. Encapsulation and Transmission

To transmit a BUM data packet, an ingress PE first pushes the ESI
label per [RFC7432] if the following conditions are all met:

o   The packet is received on a multihomed ES.

o   It's EVPN-MPLS.

o   ESI label procedure is used for split-horizon.

It then finds the S-PMSI A-D route, or the SMET/IMET route that
matches that packet. Any S-PMSI A-D route with a PTA specifying "no
tunnel information" is ignored. If one ore more SMET routes are
matched, the IMET route originated by the ingress PE for the
broadcast domain is then located to obtain the PTA.

If the found S-PMSI A-D or the IMET route has a PTA specifying
"BIER", and the ingress PE determines that BIER should be used (e.g.,
per procedures in [I-D.ietf-bess-evpn-igmp-mld-proxy] about
interworking with PEs that do not support certain tunnel types), the
(upstream-assigned) MPLS label from that PTA is pushed on the
packet's label stack in case of EVPN-MPLS. In case of EVPN-VXLAN/
NVGRE, a VXLAN/NVGRE header is prepended to the packet with the VNI/
VSID set to the value in the PTA's label field and no IP/UDP header
is used.

Then the packet is encapsulated in a BIER header and forwarded,
according to the procedures of [I-D.ietf-bier-architecture] and
[I-D.ietf-bier-mpls-encapsulation]. See especially Section 4,
"Imposing and Processing the BIER Encapsulation", of
[I-D.ietf-bier-mpls-encapsulation]. The "Proto" field in the BIER
header is set to 2 in case of EVPN-MPLS or a value to be assigned in
case of EVPN-VXLAN/NVGRE (Section 5).

In order to create the proper BIER header for a given packet, the
BFIR must know all the BFERs that need to receive that packet. If
SMET routes are matched, it determines all the BFERs from all the
matching SMET routes in the broadcast domain.

If an S-PMSI route is matched, it determines all the BFERs by finding
all the Leaf A-D routes that correspond to the S-PMSI A-D route that
is the packet's match for transmission. There are two different
cases to consider:

1 The S-PMSI A-D route that is the match for transmission carries a
   PTA that has the LIR flag set but does not have the LIR-pF flag
   set. In this case, the corresponding Leaf A-D routes are those
   whose "route key" field is identical to the NLRI of the S-PMSI A-D
   route.

2 The S-PMSI A-D route that is the match for transmission carries a
   PTA that has the LIR-pF flag. In this case, the corresponding
   Leaf A-D routes are those whose "route key" field is derived from
   the NLRI of the S-PMSI A-D route according to the procedures
   described in Section 5.2 of [EXPLICIT_TRACKING].

4.2. Disposition

The same procedures in section 3.2 of [I-D.ietf-bier-mvpn] are
followed for EVPN-MPLS (text could be copied here). For EVPN-VXLAN/
NVGRE, the only difference is that the payload is VXLAN/NVGRE and the
VNI/VSID field in the VXLAN/NVGRE header is used to determine the
corresponding mac VRF or broadcast domain.

4.2.1. At a BFER that is an Egress PE

Once the corresponding mac VRF or broadcast domain is determined from
the upstream assigned label or VNI/VSID, EVPN forwarding procedures
per [RFC7432] or [I-D.ietf-bess-evpn-overlay] are followed. In case
of EVPN-MPLS, if there is an inner label in the label stack following
the BIER header, that inner label is considered as the upstream
assigned ESI label for split horizon purpose.

4.2.2. At a BFER that is a P-tunnel Segmentation Boundary

This is only applicable to EVPN-MPLS. The same procedures in
Section 3.2.2 of [I-D.ietf-bier-mvpn] are followed, subject to
multihoming considerations described in Section 3 of this document.

5. IANA Considerations

This document requests two assignments in "BIER Next Protocol
Identifiers" registry, with the following two recommended values:

o 7: Payload is VXLAN encapsulated (no IP/UDP header)

o 8: Payload is NVGRE encapsulated (no IP header)

6. Security Considerations

To be updated.

7. Acknowledgements

The authors thank Eric Rosen for his review and suggestions.
Additionally, much of the text is borrowed verbatim from
[I-D.ietf-bier-mvpn].

8. References

8.1. Normative References

[I-D.ietf-bess-evpn-bum-procedure-updates]
          Zhang, Z., Lin, W., Rabadan, J., and K. Patel, "Updates on
          EVPN BUM Procedures", draft-ietf-bess-evpn-bum-procedure-
          updates-01 (work in progress), December 2016.

[I-D.ietf-bess-evpn-igmp-mld-proxy]
          Sajassi, A., Thoria, S., Patel, K., Yeung, D., Drake, J.,
          and W. Lin, "IGMP and MLD Proxy for EVPN", draft-ietf-
          bess-evpn-igmp-mld-proxy-00 (work in progress), March
          2017.

[I-D.ietf-bess-mvpn-expl-track]
          Dolganow, A., Kotalwar, J., Rosen, E., and Z. Zhang,
          "Explicit Tracking with Wild Card Routes in Multicast
          VPN", draft-ietf-bess-mvpn-expl-track-02 (work in
          progress), June 2017.

   [I-D.ietf-bier-architecture]
             Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and
             S. Aldrin, "Multicast using Bit Index Explicit
             Replication", draft-ietf-bier-architecture-06 (work in
             progress), April 2017.

   [I-D.ietf-bier-mpls-encapsulation]
             Wijnands, I., Rosen, E., Dolganow, A., Tantsura, J.,
             Aldrin, S., and I. Meilik, "Encapsulation for Bit Index
             Explicit Replication in MPLS and non-MPLS Networks",
             draft-ietf-bier-mpls-encapsulation-07 (work in progress),
             June 2017.

   [I-D.ietf-bier-mvpn]
             Rosen, E., Sivakumar, M., Aldrin, S., Dolganow, A., and T.
             Przygienda, "Multicast VPN Using BIER", draft-ietf-bier-
             mvpn-05 (work in progress), January 2017.

   [I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements]
             Zhang, Z., Kebler, R., Lin, W., and E. Rosen, "MVPN/EVPN
             C-Multicast Routes Enhancements", draft-zzhang-bess-mvpn-
             evpn-cmcast-enhancements-00 (work in progress), July 2016.

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
             Requirement Levels", BCP 14, RFC 2119,
             DOI 10.17487/RFC2119, March 1997,
             <http://www.rfc-editor.org/info/rfc2119>.

   [RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A.,
             Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based
             Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February
             2015, <http://www.rfc-editor.org/info/rfc7432>.

8.2. Informative References

   [I-D.ietf-bess-evpn-overlay]
             Sajassi, A., Drake, J., Bitar, N., Shekhar, R., Uttaro,
             J., and W. Henderickx, "A Network Virtualization Overlay
             Solution using EVPN", draft-ietf-bess-evpn-overlay-08
             (work in progress), March 2017.

Authors' Addresses

   Zhaohui Zhang
   Juniper Networks

   EMail: zzhang@juniper.net

   Antoni Przygienda
   Juniper Networks

   EMail: prz@juniper.net

   Ali Sajassi
   Cisco Systems

   EMail: sajassi@cisco.com

   Jorge Rabadan
   Nokia

   EMail: jorge.rabadan@nokia.com