Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 15, 2019

J. Xie
Huawei Technologies
L. Geng
L. Wang
China Mobile
M. McBride
G. Yan
Huawei

August 14, 2018

Segmented MVPN Using IP Lookup for BIER
draft-xie-bier-mvpn-segmented-02

Abstract

This document specifies an alternative of the control plane and data plane procedures that allow segmented MVPN using BIER. It allows the use of the more efficient LIR-pF explicit-tracking for segmented MVPN deployment when BIER is used as the upstream or downstream segments. It requires a segmentation point BFR doing an IP header lookup, which is common for the forwarding procedure on BFER, or the forwarding procedure on ABR with local VPN CEs connected. This document updates [I-D.ietf-bier-mvpn].

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].

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 February 15, 2019.

Table of Contents

1. Introduction

When using BIER to transport an MVPN data packet through a BIER

domain, an ingress PE functions as a BFIR (see [RFC8279]). The BFIR must determine the set of BFERs to which the packet needs to be delivered. This can be done through an explicit-tracking function using a LIR and/or LIR-pF flag in BGP MVPN routes, per the

[RFC6513],[RFC6514],[RFC6625],[I-D.ietf-bess-mvpn-expl-track], and [I-D.ietf-bier-mvpn].

Using a LIR-pF Flag will bring some extra benefits, as [I-D.ietf-bier-mvpn] and [I-D.ietf-bess-mvpn-expl-track] have stated. But unfortunately, the LIR-pF explicit tracking for a segmented MVPN deployment is not allowed in the current draft [I-D.ietf-bier-mvpn], because the draft requires a per-flow upstream-assigned label to do the data-plane per-flow lookup on the segmentation point BFR.

This document specifies an alternative of the control plane and data plane procedures that allow segmented MVPN using BIER in both segments. This allows the use of the more efficient LIR-pF explicit-tracking as the BIER overlay, with a slight change in the forwarding procedure of a segmentation point BFR by using IP lookup. This will bring some significant benefits to the segmented MVPN deployment, including:

- Getting a much better multicast join latency by eliminating the round trip interaction of S-PMSI AD routes and Leaf AD routes. Especially, the S-PMSI A-D routes may need a data-driven procedure to trigger, and make the multicast join latency even worse.

- Greatly reducing the number of S-PMSI A-D routes that BFIR and BFERs need to save.

- Consolidated forwarding procedure of IP lookup for every BIER Overlay functioning routers, such as BFIR, BFER, segmentation point BFR, and segmentation point BFR with BFER function.

2. Terminology

Readers of this document are assumed to be familiar with the terminology and concepts of the documents listed as Normative References.

3. Problem Statement and Considerations

3.1. Problem Statement

BIER is a stateless multicast forwarding by introducing a multicast-specific BIER header in the data plane. The maximal number of BFERs a packet can reach is limited by the bit string length of a BIER header. For a network with many routers in multiple IGP areas (typically an Inter-Area network), it may be more expected to use a segmented MVPN when deploying BIER than traditional MVPN.

However, it is not allowed in the [I-D.ietf-bier-mvpn] to use a LIR-pF explicit-tracking when deploying a segmented MVPN. This will lead to a low efficiency of explicit-tracking, and cause a worse multicast join latency. Here we take a scenario of inter-area segmented MVPN with both segments using BIER as an example.

3.2. Considerations

A BFIR is always needed to know the BFERs interested in a specific flow. This is a function of a BIER overlay defined in [RFC8279]. A segmentation point BFR in a segmented MVPN deployment, saying ABR, will play similar roles of both BFIR and BFER. It needs to do a disposition of a BIER Header, and then do an imposition of a new BIER Header. It requires the ABR router to maintain per-flow states, and especially, such per-flow states always include a set of BFERs who are intrested in a specific flow by using an explicit-tracking procedure.

This behavior is completely different from a traditional segmented MVPN deployment, e.g, with both of the two segments using P2MP label switch.

In a traditional segmented MVPN with both segments using P2MP label switch, it is expected to receive a MPLS packet and replicate to downstream routers after swap the MPLS Label. A lookup of IP packet is not expected. Also, in a traditional segmented MVPN deployment, an MPLS label represents a P-tunnel, which may carry one, many or even all multicast flow(s) of a VPN, so it is not always a per-flow state on the segmentation point router.

In conclusion, the pattern of forwarding packets on segmentation points only by lookup of MPLS label mapped from multicast flow(s) is significantly unnecessary when BIER is introduced. Instead, doing a per-flow lookup of IP header on segmentation points is more efficient and consolidated.

4. Upstream BIER and Downstream BIER

4.1. Explicit Tracking using LIR-pF

In a scenario of Inter-area Segmented MVPN with both segments using BIER, the determination of the set of BFERs that need to receive a specific multicast flow of (C-S1,C-G1) in each segment, can be obtained by using a LIR-pF flag.

Suppose a topology of this:

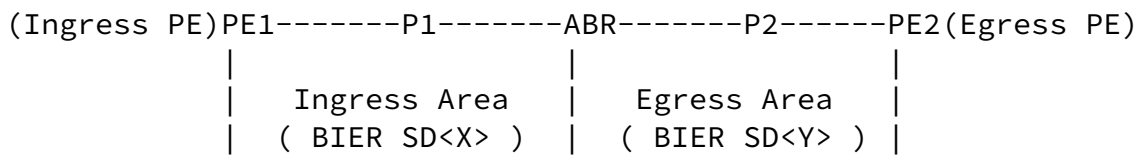

Figure 1: Example topology

PE1 is Ingress PE, and the area of { PE1 -- P1 -- ABR } is called an Ingress Area.

PE2 is Egress PE, and the area of { ABR -- P2 -- PE2 } is called an Egress Area.

The Ingress PE is configured to use a BIER tunnel type for a MVPN instance for the Ingress Area, and the ABR is configured to use a BIER tunnel type for the MVPN instance for the Egress Area.

The Ingress PE originates a wildcard S-PMSI A-D route (C-*,C-*) and the PTA of that route has the following settings:

- The LIR-pF and LIR flags be set.

- The tunnel type be set to "BIER".

- A non-zero MPLS label be specified.

ABR receives the S-PMSI A-D route from the Ingress PE, and re-advertises the route to the Egress PE, with a PTA type "BIER", and PTA flags of LIR and LIR-pF, and a new non-zero upstream-assigned MPLS label allocated by ABR per-VPN.

Egress PE receives the S-PMSI A-D route from the ABR, and checks if it need to response with a Leaf A-D route to this S-PMSI A-D route using the process of the "match for reception" and "match for tracking" as defined in [I-D.bess-mvpn-expl-track]. In this example, for a C-flow of (C-S1, C-G1), the checking result of "matched for tracking" is the S-PMSI(C-*, C-*), and the checking result of "matched for reception" is also the S-PMSI(C-*, C-*). Egress PE will then send a Leaf A-D route (RD, C-S1, C-G1, Root=PE1, Leaf=PE2) to the ABR with a PTA flag LIR-pF, and a Leaf A-D route (RD, C-*, C-*, Root=PE1, Leaf=PE2) without a PTA flag LIR-pF.

ABR then has an explicit-tracking result of a new per-flow information of (RD, C-S1, C-G1, Root=PE1) with Egress PE as its leaf or BFER. ABR's "matched for tracking" result to this flow(RD, C-S1, C-G1, PE1) will then be updated with a new record, and ABR then sends a Leaf A-D route (RD, C-S1, C-G1, Root=PE1, Leaf=ABR) to Ingress PE.

Ingress PE then has an explicit-tracking result of a new per-flow information of (RD, C-S1, C-G1, Root=PE1) with ABR as its leaf or BFER.

From this procedure description one can see that:

1. The S-PMSI A-D(C-*, C-*) route is functioning as a per-VPN anchor of the upstream and the downstream(s), which can be called an MVPN FEC in this document, saying MVPN FEC(*,*).

2. The Leaf A-D(S,G) routes are functioning as a per-flow anchor of the downstream(s) and the upstream, which are also MVPN FECs accordingly, saying MVPN FEC(S,G).

3. The Tuple of (Root=PE1, RD) in S-PMSI (C-*, C-*) or Leaf AD(C-*,

C-*) or Leaf AD(C-S, C-G) represents an VRF on the ABR implicitly.

ABR knows the per-vpn infmation of a (Root=PE1, RD) tuple when receiving and re-advertising the S-PMSI A-D(*,*) route bound with a PTA, where:

- Inbound SD (InSD): in PTA of the received S-PMSI(*,*) route.
- Inbound VpnLabel (InVpnLabel): in PTA of the received S-PMSI(*,*) route.
- Inbound BfirId (InBfirId): in PTA of the received S-PMSI(*,*) route.
- Outbound SD(OutSD): in PTA of the re-advertised S-PMSI(*,*) route.
- Outbound VpnLabel (OutVpnLabel): in PTA of the re-advertised S-PMSI(*,*) route.
- Outbound BfirId (OutBfirId): in PTA of the re-advertised S-PMSI(*,*) route.

ABR establishs a per-flow control-plane state accordingly like this:

- Per-flow upstream state, according to the Leaf A-D (C-S, C-G) route send to the Ingress PE: (PE1, RD, C-S1, C-G1, InSD, InBfirId, InVpnLabel).
- Per-flow downstream state(s), according to the Leaf A-D(C-S, C-G) route(s) received by the ABR from Egress PE(s): (PE1, RD, C-S1, C-G1, Leaf, OutSD, OutBfirId, OutVpnLabel).

ABR knows the BIER Label(s) it allocated for InSD and OutSD, saying InBierLabel for InSD<X> and OutBierLabel for OutSD<Y>, and thus it can establish the per-flow forwarding state:

- Per-flow upstream forwarding state: (InBierLabel, InBfirId, InVpnLabel, C-S1, C-G1).
- Per-flow downstream(s) forwarding state: (InBierLabel, InBfirId,

InVpnLabel, C-S1, C-G1, Leaf, OutBfirId, OutVpnLabel, OutBitString)

4.2. Forwarding using IP lookup on Segmentation Point

The Forwarding procedure of a segmentation point, ABR, have the following steps:

1. Need to do a disposition of BIER encapsulation, and a upstream-assigned VpnLabel lookup in the special context to get the appropriate VPN represented by Ingress PE's IP and RD.

2. Need to do a imposition of BIER by an IP lookup to get the appropriate BIER encapsulation for Area<Y>.

3. Optionally the same disposition of BIER encapsulation and IP lookup are also required on ABR if ABR has local VPN CEs.

For Step 1 and step 2, One can think them as one-to-many swapping of a big 'BIER header' instead of a small 'Label' in P2MP case:

- swapping the InBierLabel with an OutBierLabel.

- swapping the InBfirId with an OutBfirId.

- swapping the InVpnLabel with an OutVpnLabel.

- swapping the InBitString with an OutBitString.

The key of a per-flow lookup on ABR is a tuple of (InBierLabel, InBfirId, InVpnLabel) and a tuple of (C-S1, C-G1), representing a VRF and a flow respectively. All the elements are from a BIER packet, and such an IP lookup can be seen the same as an MFIB lookup, if the (InBierLabel, InBfirId, InVpnLabel) tuple is mapped to a VRF locally on the ABR.

5. Upstream P2MP/IR and Downstream BIER

5.1. Explicit Tracking using LIR-pF

The procedures described in chapter 4.1 is also suitable for this case, provided the LIR-pF explicit-tracking is used appropriately.

Suppose a topology of this:

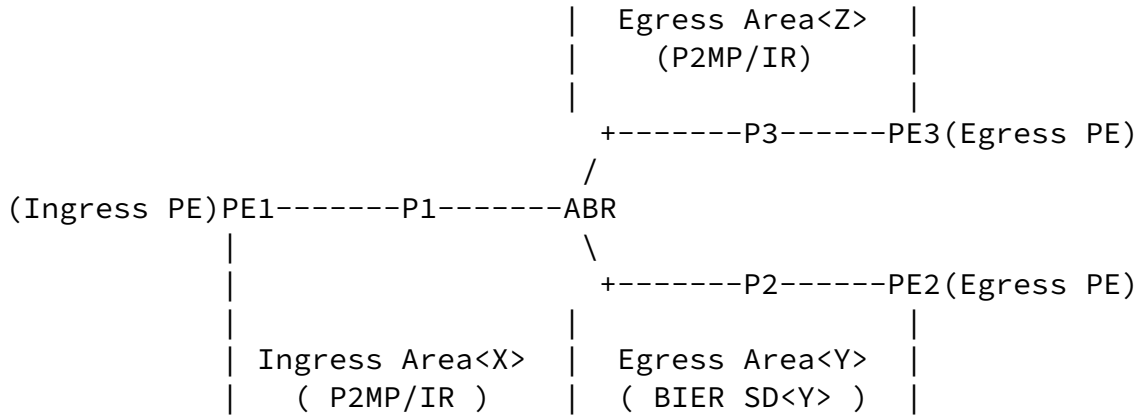

Figure 2: Example topology

The Ingress PE is configured to use a P2MP or IR tunnel type for a MVPN instance for the Ingress Area<X>, and the ABR is configured to use a BIER tunnel type for the MVPN instance for the Egress Area <Y>, and ABR may be configured to use a P2MP or IR tunnel type for another Egress Area<Z>.

Example 1: Use Inclusive P-tunnel for traditional areas.

PE1 may configure to use one SPMSI (*,*,PTA<mLDP, Flag=LIR+LIRpF>) route , for one Unidirectional Inclusive mLDP P2MP tunnel.

ABR may configure to reflect only the SPMSI(*,*) route with the PTA type changed to BIER for Area<Y>, and reflect the SPMSI(*,*) route with the PTA type changed to mLDP, RSVP-TE or IR for Area<Z>.

Example 2: Use Selective P-tunnel for traditional areas.

PE1 may configure to use one SPMSI (*,*,PTA<mLDP, Flag=LIR+LIRpF>) route and one to many SPMSI (S,G,PTA<mLDP, Flag=LIR+LIFpF>) routes, for one Unidirectional Inclusive mLDP P2MP tunnel and one to many Selective mLDP P2MP tunnels respectively.

ABR may configure to reflect only the SPMSI(*,*) route with the PTA type changed to BIER for Area<Y>, and reflect the SPMSI(*,*) route

and SPMSI(S,G) routes with the PTA type changed to mLDP, RSVP-TE or IR for Area<Z>.

5.2. Forwarding using IP lookup on Segmentation Point

The Forwarding procedure of a segmentation point, ABR, have 3 conditions:

1. Need to do an P2MP MPLS switching for Area<Z>, and this does not change.

2. Need to do a disposition of P2MP or IR tunnel, and a following IP lookup to get the appropriate BIER encapsulation for Area<Y>.

3. Need to do the same IP lookup to get the appropriate local VRF interfaces if ABR has local VPN CEs.

6. Upstream BIER and Downstream P2MP/IR

6.1. Explicit Tracking using LIR-pF

The procedures described in chapter 4.1 is also suitable for this case, provided the LIR-pF explicit-tracking is used appropriately.

Suppose a topology of this:

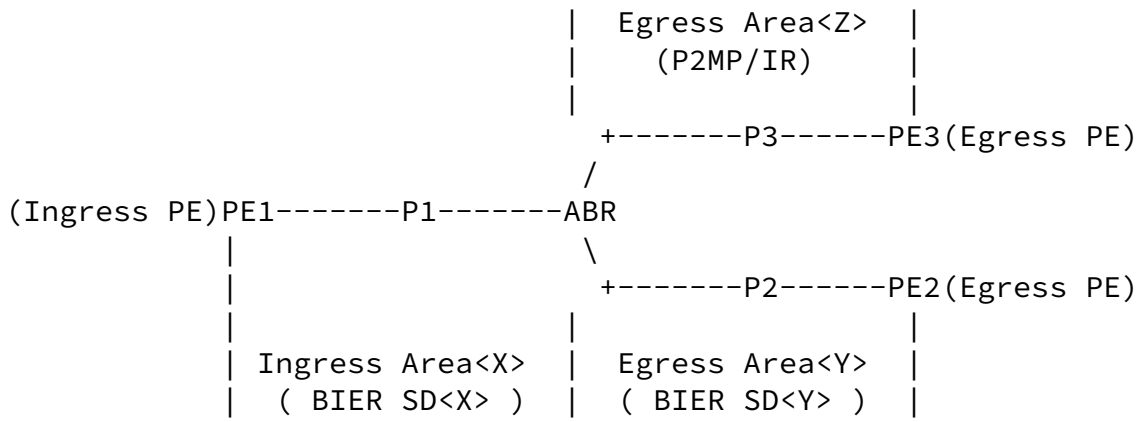

Figure 3: Example topology

The Ingress PE is configured to use a BIER tunnel type for a MVPN instance for the Ingress Area<X>, and the ABR is configured to use a BIER tunnel type for the MVPN instance for the Egress Area <Y>, and ABR may be configured to use a P2MP or IR tunnel type for another Egress Area<Z>.

Example 1: Use Inclusive P-tunnel for traditional areas.

PE1 may configure to use one SPMSI (*,*,PTA<BIER, Flag=LIR+LIRpF>) route , for one BIER tunnel.

ABR may configure to reflect only the SPMSI(*,*) route with the PTA type unchanged for Area<Y>, and reflect the SPMSI(*,*) route with the PTA type changed to mLDP, RSVP-TE or IR for Area<Z>.

Example 2: Use Selective P-tunnel for traditional areas.

PE1 may configure to use one SPMSI (*,*,PTA<BIER, Flag=LIR+LIRpF>) route and one to many SPMSI (S,G,PTA<BIER, Flag=LIR+LIFpF>) routes. They include the same one BIER tunnel, but the many SPMSI(S,G) routes can initialize the many Inter-area MVPN FECs, and thus enable the ABR to initialize many SPMSI tunnels for Area<Z>. The ABR need to support allocating SPMSI tunnels for Area<Z> per the upstream SD<X> and the (S,G).

ABR may configure to reflect only the SPMSI(*,*) route with the PTA type changed to BIER for Area<Y>, and reflect the SPMSI(*,*) route and SPMSI(S,G) routes with the PTA type changed to mLDP, RSVP-TE or IR for Area<Z>.

6.2. Forwarding using IP lookup on Segmentation Point

The Forwarding procedure of a segmentation point, ABR, have the following conditions:

1. Need to do a disposition of BIER encapsulation, and an upstream-assigned VpnLabel lookup in the special context to get the appropriate VPN represented by Ingress PE's IP and RD.

2. Need to do an imposition of BIER by an IP lookup to get the appropriate BIER encapsulation for Area<Y>.

3. Need to do an imposition of P2MP by an IP lookup to get the appropriate P2MP downstreams.

4. Optionally the same disposition of BIER encapsulation and IP lookup are also required on ABR if ABR has local VPN CEs.

7. Summary and Recommendations

Traditional P2MP forwarding on an router can be simple if it is not the Bud router case. Only a one shot Label lookup is enough to get the outgoing P2MP interfaces to replicate. The same benefit is also available by a pure 'ABR' node for segmented MVPN scenario.

But for a Bud router, the forwarding of an MPLS P2MP packet must include a label lookup to find the outgoing P2MP interfaces, and a further IP lookup for the local overlay forwarding. The same challenge is same to a ABR node who has local VPN CEs connected.

For a segmented MVPN deployment with BIER, there are much more ingredients different from the simple P2MP label lookup and swap. This is because the BIER label is indicating a BIER layer aggregrated tunnel for all flows, while the further disposition or imposition have to do an overlay lookup. The benefits of insisting on Label lookup, including the per-flow upstream-assigned VpnLabel lookup, on a boundary router forwarding procedure is relatively low when BIER is used on the upstream segment or downstream segments of the boundary router, unless the boundary router is unable to do such an IP lookup. This capability of IP lookup after BIER disposition is basic for BFER handling, and for case when the ABR has local VPN CEs connected.

Recommendations are:

1. that implementations support the IP lookup for Segmented BIER MVPN if it support BFER function which require a disposition of BIER header and a further IP lookup.

2. that implementations support the LIR-pF explicit tracking for Segmented BIER MVPN if it support LIR-pF explicit tracking for intra-area MVPN.

3. that implementations support the IP lookup and LIR-pF explicit tracking for Segmented BIER MVPN if one want to gain better multicast join latency and flood less SPMSI(S,G) routes to every ABRs and PEs in segmented MVPN deployment.

8. Security Considerations

The procedures of this document do not, in themselves, provide privacy, integrity, or authentication for the control plane or the data plane.

9. IANA Considerations

No IANA allocation is required.

10. Acknowledgements

TBD.

11. References

11.1. Normative References

[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-09 (work in progress), April 2018.

[I-D.ietf-bier-mvpn]
Rosen, E., Sivakumar, M., Aldrin, S., Dolganow, A., and T. Przygienda, "Multicast VPN Using BIER", draft-ietf-bier-mvpn-11 (work in progress), March 2018.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

[RFC6625] Rosen, E., Ed., Rekhter, Y., Ed., Hendrickx, W., and R.

Qiu, "Wildcards in Multicast VPN Auto-Discovery Routes", RFC 6625, DOI 10.17487/RFC6625, May 2012, <https://www.rfc-editor.org/info/rfc6625>.

[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>.

11.2. Informative 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>.

Authors' Addresses

Jingrong Xie
Huawei Technologies

Email: xiejingrong@huawei.com

Liang Geng
China Mobile
Beijing 10053

Email: gengliang@chinamobile.com

Lei Wang
China Mobile
Beijing 10053

Email: wangleiyjy@chinamobile.com

Mike McBride
Huawei

Email: mmcbride7@gmail.com

Gang Yan
Huawei

Email: yangang@huawei.com