BESS
Internet-Draft
Updates: 6513, 6514, 7524 (if approved)
Intended status: Standards Track
Expires: June 23, 2019

Z. Zhang
Juniper Networks
J. Xie
Huawei

December 20, 2018

MVPN/EVPN Segmentated Forwarding Options
draft-zzhang-bess-mvpn-evpn-segmented-forwarding-00

Abstract

[RFC6513] and [RFC6514] specify MVPN Inter-AS Segmentation procedures. [RFC7524] specifies MVPN Inter-Area Segmentation procedures. [I-D.ietf-bess-evpn-bum-procedure-updates] specifies EVPN BUM Inter-Region Segmentation Procedures. Several other documents also touch upon the segmentation topic. The forwarding at the segmentation points has been assumed to be label switching, subject to certain limitations. The purpose of this document is to provide a review of segmentation points' available forwarding options and limitations, and to clarify and expand some procedures.

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 June 23, 2019.

Table of Contents

1. Terminology 2
2. Introduction 3
2.1. MPLS Label Switching at Segmentation Points 3
2.2. IP Processing at Segmentation Points 5
3. Specifications 6
4. References 6
4.1. Normative References 6
4.2. Informative References 7
Authors' Addresses 8

1. Terminology

This document uses terminology from MVPN and EVPN. It is expected that the audience is familiar with the concepts and procedures defined in [RFC6513], [RFC6514], [RFC7524], [RFC7432], [I-D.ietf-bess-evpn-bum-procedure-updates], and [I-D.ietf-bess-evpn-igmp-mld-proxy]. Some terms are listed below for references.

- PMSI: P-Multicast Service Interface - a conceptual interface for a PE to send customer multicast traffic to all or some PEs in the same VPN. A PMSI A-D route is a BGP MVPN/EVPN auto-discovery route that announces the PMSI and optionally the tunnel that instantiates the PMSI.

- I-PMSI: Inclusive PMSI - to all PEs in the same VPN.

- S-PMSI: Selective PMSI - to some of the PEs in the same VPN.

- Leaf A-D routes: For explicit leaf tracking purpose. Triggered by S-PMSI A-D routes and targeted at triggering route's originator.

- IMET A-D route: Inclusive Multicast Ethernet Tag A-D route. The EVPN equivalent of MVPN Intra-AS I-PMSI A-D route.

- SMET A-D route: Selective Multicast Ethernet Tag A-D route. The EVPN equivalent of MVPN Leaf A-D route but unsolicited and untargeted.

2. Introduction

[RFC6513] and [RFC6514] specify MVPN Inter-AS Segmentation procedures. [RFC7524] specifies MVPN Inter-Area Segmentation procedures. [I-D.ietf-bess-evpn-bum-procedure-updates] specifies EVPN BUM Inter-Region Segmentation Procedures. Several other documents also touch upon the segmentation topic.

2.1. MPLS Label Switching at Segmentation Points

It has been assumed that the forwarding across a segmentation point is label based. The upstream segment of a PMSI tunnel is stitched to the downstream segment via label switching and no IP processing is done. This is true even if the segmentation point also has a VRF with PE-CE interfaces, where IP processing is done to decide if a packet should be forwarded out of a PE-CE interface but label switching is used for forwarding traffic to receivers connected by downstream segments.

This label switching is based on the assumption/requirement that each PMSI tunnel has its own unique label (in the simpliest case - this can be relaxed as specified in [RFC7988] in case of Ingress Relication). The following is a breakdown of the various situations:

- If an aggregated RSVP-TE or mLDP P2MP tunnel, or BIER is used for the upstream (or downstream) segment, the x-PMSI A-D route received (or re-advertised, in case of downstream segment) by the segmentation point carries a per-PMSI label in the PMSI Tunnel Attribute (PTA). The BIER case is specified in [I-D.ietf-bier-mvpn] and [I-D.ietf-bier-evpn].

- If a unique RSVP-TE or mLDP P2MP tunnel is used for for each upstream segment, the segmentation point advertises a unique label for each tunnel to the upstream node on the tunnel. Similarly, in the downstream segment case, the segmenation point must receive a unique tunnel label.

- If Ingress Replication is used for the upstream segment, the segmentatation point may either simply advertise a different label in each Leaf A-D route that it advertises, or use a more elaborate procedure to decide how labels could be advertised while still

allow correct label switching procedure, as specified in Section 7.2 of [RFC7988].

Notice that in the case of P2MP tunnel, x-PMSI A-D routes are required to advertise the tunnel identification and in case of tunnel aggregation (BIER or aggregated P2MP tunnel) the x-PMSI A-D routes are required to advertise the per-PMSI label. However, [I-D.ietf-bess-mvpn-expl-track] introduces a "Leaf Information Required per Flow" bit (LIR-pF) in the flags field of the PTA of wildcard S-PMSI A-D routes, so that an ingress PE does not have to advertise individual more specific S-PMSI A-D routes even if it wants to explicitly track the leaves for more specific flows. This can be used for RSVP-TE P2MP, Ingress Replication and BIER.

For EVPN, explicit tracking is based on unsolicited Selective Multicast Ethernet Tag (SMET) A-D routes and LIR-pF is not used. However, that is as if the LIR-pF flag was set in an implicit (C-*, C-*) wildcard S-PMSI A-D route.

Both [I-D.ietf-bier-mvpn] and [I-D.ietf-bier-evpn] specify that the LIR-pF flag MUST not be used with segmentation. That’s because with LIR-pF while an ingress PE can send a flow to only leaves tracked for the flow, it does not advertise the label bound to the corresponding PMSI for the flow (as the LIR-pF removes the need to advertise the more specific S-PMSI routes).

The same restriction also applies if aggregated RSVP-TE P2MP tunnels are used (the same tunnel could be used for multiple more specific S-PMSIs but a per-PMSI label would be associated with each S-PMSI). The LIR-pF flag removes the need for those more specific S-PMSI A-D routes so no S-PMSI specific labels could be advertised for the segmentation points to do label switching with.

The restriction does not apply to Ingress Replication because the per-PMSI label is advertised in the Leaf A-D routes.

The restriction with BIER and aggregated RSVP-TE P2MP tunnel can be lifted if the LIR-pF triggered more specific MVPN Leaf A-D routes or the unsolicited EVPN SMET routes can trigger corresponding S-PMSI A-D routes, so that the per-PMSI labels can be advertised. The concept of triggering S-PMSI A-D routes by Leaf/SMET A-D routes is already present in [RFC7524] and [I-D.zzhang-bess-mvpn-evpn-cmcast-enhancements].

It may be argued that triggering S-PMSI A-D route from Leaf/SMET A-D routes for mroe specific flows has the following concerns (which leads to the consideration for forwarding option described in Section 2.2):

- Flooding of those extra more specific S-PMSI A-D routes

- Delay in setting up the forwarding state (as the segmentation points now have to wait for the corresponding S-PMSI A-D route from its upstream).

The first concern can be discounted that the burden of those extra S-PMSI A-D routes are mainly in the control plane. The forwarding plane does need to maintain addtional per-PMSI labels but it's much better than the alertnative described in the following section.

The second concern can be mitigated by having the ingress PE delay switching traffic over to the more specific S-PSMI. That way, traffic will continue to be forwarded on the less specific PMSI (and label switched by segmentation points) for a short period before being moved to the more specific S-PMSI.

2.2. IP Processing at Segmentation Points

If the above mentioned discount/mitigation are not enough to address the two concerns, IP processing can be used at segmentation points. This will allow the use of LIR-pF with segemntation without triggering those more specific S-PMSI A-D routes [I-D.xie-bier-mvpn-segmented] .

Basically, a segmentation point will create an IP multicast forwarding table for each "context", which could be for an EVPN Broadcast Domain (BD), a L3 VPN, an L3 VPN Extranet, or even something of smaller scope. An incoming packet on an upstream segment is decapsulated and a corresponding IP multicat forwarding table is identified. An IP lookup is performed and forwarded into downstream segments accordingly.

While this does not require the S-PMSI A-D routes triggered by Leaf/ SMET routes (and corresponding label forwarding state), additional IP forwarding tables and lookup are needed, which requires additional memory and cycles in the forwarding path, additional code to maintain the RIB/FIB tables, and additional OPEX to monitor them.

Nonetheless, if IP processing on a segmentation point is desired for the reason of LIR-pF bit, the following could be done.

- Wildcard S-PMSI A-D routes with the LIR-pF flag are assigned with different labels from those in x-PMSI routes w/o the flag, and they lead to IP lookup. The labels can either be upstream assigned or assigned from a Domain-wide Common Block (DCB) [I-D.ietf-bess-mvpn-evpn-aggregation-label].

- Labels in x-PMSI routes w/o the LIR-pF flag, which are different from those in routes with the flag, lead to label switching.

- A Leaf A-D route with LIR-pF flag triggers corresponding (C-S, C-G) or (C-*, C-G) routes used for IP lookup, if there is no corresponding S-PMSI A-D route with LIR-pF flag.

- Upstream PE/ABR uses the label advertised in the matching x-PMSI routes to send traffic (so the packets will either be label switched or ip forwarded by segmentation points).

On a PE, there are already VRFs or BDs configured so the IP RIBs/FIBs are just in those VRFs/BDs. On a segmentation point, most likely there are no VRFs/BDs. How IP RIBs/FIBs are managed is local behavior and implementation dependent. While it is outside the scope of this document, one method could be to maintain one IP RIB/FIB for each label carried in a wildcard S-PMSI A-D route with the LIR-pF flag. .

3. Specifications

Detail specification for the above summary will be added in upcoming revisions.

4. References

4.1. Normative References

[I-D.ietf-bess-evpn-bum-procedure-updates]
Zhang, Z., Lin, W., Rabadan, J., Patel, K., and A. Sajassi, "Updates on EVPN BUM Procedures", draft-ietf-bess-evpn-bum-procedure-updates-05 (work in progress), December 2018.

[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-02 (work in progress), June 2018.

[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-13 (work in progress), November 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>.

[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>.

[RFC7524] Rekhter, Y., Rosen, E., Aggarwal, R., Morin, T., Grosclaude, I., Leymann, N., and S. Saad, "Inter-Area Point-to-Multipoint (P2MP) Segmented Label Switched Paths (LSPs)", RFC 7524, DOI 10.17487/RFC7524, May 2015, <https://www.rfc-editor.org/info/rfc7524>.

4.2. Informative References

[I-D.ietf-bess-mvpn-evpn-aggregation-label]
Zhang, Z., Rosen, E., Lin, W., Li, Z., and I. Wijnands, "MVPN/EVPN Tunnel Aggregation with Common Labels", draft-ietf-bess-mvpn-evpn-aggregation-label-02 (work in progress), December 2018.

[I-D.ietf-bier-evpn]
Zhang, Z., Przygienda, T., Sajassi, A., and J. Rabadan, "EVPN BUM Using BIER", draft-ietf-bier-evpn-01 (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.

[I-D.xie-bier-mvpn-segmented]
Xie, J., Geng, L., Wang, L., McBride, M., and G. Yan, "Segmented MVPN Using IP Lookup for BIER", draft-xie-bier-mvpn-segmented-06 (work in progress), October 2018.

[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.

[RFC7988] Rosen, E., Ed., Subramanian, K., and Z. Zhang, "Ingress Replication Tunnels in Multicast VPN", RFC 7988, DOI 10.17487/RFC7988, October 2016, <https://www.rfc-editor.org/info/rfc7988>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks

EMail: zzhang@juniper.net

Jingrong Xie
Huawei

EMail: xiejingrong@huawei.com