Workgroup: bess
Internet-Draft:
draft-zzhang-mvpn-evpn-controller-01
Published: 25 October 2021
Intended Status: Standards Track
Expires: 28 April 2022
Authors: Z. Zhang
Juniper Networks
R. Parekh
Cisco Systems
Z. Zhang
ZTE
H. Bidgoli
Nokia

MVPN and EVPN BUM Signaling with Controllers

Abstract

This document specifies optional procedures for BGP-MVPN and EVPN BUM signaling with controllers. When P2MP tunnels used for BGP-MVPN and EVPN BUM are to be signaled from controllers, the controllers can learn tunnel information (identifier, root, leaf) by participating BGP-MVPN and EVPN BUM signaling, instead of relying on ingress PEs to collect the information and then pass to the controllers. Additionally, Inclusive/Selective PMSI Auto Discovery Routes can be originated from controllers based on central provisioning, instead of from PEs based on local provisioning.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

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 28 April 2022.

Table of Contents

1. Terminologies

Familiarity with MVPN/EVPN protocols and procedures is assumed. Some terminologies are listed below for convenience.

*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/BD.

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

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

*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.

As pointed out above, the EVPN IMET route is the equivalent of MVPN I-PMSI A-D route. In the rest of the document, unless explicitly stated, I-PMSI A-D route refers to MVPN Intra-AS I-PMSI A-D route and/or EVPN IMET route.

2. Introduction

Consider a provider network with BGP-MVPN/EVPN where controllers are used to set up P2MP tunnels per [I-D.ietf-bess-bgp-multicast-controller] or [I-D.ietf-pim-sr-p2mp-policy]. For a controller to calculate the corresponding trees and set up the tunnels, it needs to learn the (ID, root, leaf) information for those trees. Currently, [I-D.ietf-bess-mvpn-evpn-sr-p2mp] specifies that an ingress PE assigns the SR P2MP ID and collects leaf information via Leaf A-D routes, and then pass onto the controller. Observing that BGP-MVPN/EVPN signaling typically involves Router Reflectors, which may typically be hosted on or co-located with controllers, it makes sense to have the controllers participating BGP-MVPN/EVPN signaling to learn (ID, root, leaf) information. This will relieve the PEs from maintaining Leaf A-D routes, and remove the extra hop of leaf information propagation.

Also Consider that in the same network many selective tunnels are used, and their usages are dynamically provisioned based on specific needs at different time. For example, the provider provides video transmission services for events at various time, location and to various receivers. With traditional methods the provider would provision the PEs at the transmission sources with various selective tunnels, which triggers corresponding S-PMSI A-D routes. The provisioning is put in place shortly before an event takes place and removed shortly after the event ends. Alternatively and preferrably, a controller can originate S-PMSI A-D routes based on centralized provisioning on behalf of the source PEs. The controller also collects the leaf information (either based on centralized provisioning or based on Leaf A-D routes), calculates the tree and signal tree nodes. Additionally, when tunnel aggregation labels are allocated from Domain-wide Common Block (DCB), originating I/S-PMSI A-D routes from controllers makes the DCB label allocation a lot easier.

It is possible that an operator prefers automatic DCB aggregation label allocation by the controller but prefers I/S-PMSI A-D routes origination from individual PEs. In that case, a PE can target an I/S-PMSI A-D route at the controller and the controller will allocate a DCB label and return it in a corresponding Leaf A-D route.

3. Specification

The procedures specified in this section applies if one or more controllers participate MVPN/EVPN signaling for the purpose of leaf discovery for P2MP tree calculation, and/or if controllers are to originate I/S-PMSI A-D routes or BGP-MVPN and/or BGP-EVPN BUM.

3.1. Controller Address Extended Community

This document defines a new Transitive IPv4-Address-Specific Extended Community Sub-Type: "Controller Address". This document also defines a new BGP Transitive IPv6-Address-Specific Extended Community Sub-Type: "Controller Address".

A Controller Address Extended Community (referred to as Controller EC) is constructed by setting the Global Administrator field to the IP address of the controller and the Local Administrator field to 0.

3.2. Targeting Leaf A-D Routes to Controllers

When a PE originates an I/S-PMSI A-D route with PTA's tunnel type set to PIM-SSM/ASM, mLDP or SR P2MP that are to be set up by controllers, the PE MUST attach a Controller EC constructed as above. If there are multiple controllers, then one Controller EC is attached for each of the controllers.

In case of tunnel segmentation and a new controller is used for the next segmentation region, when an ABR/ASBR/RBR re-advertises the I/S-PMSI A-D route to the next segmentation region it MUST modify the Controller EC to specify the new controller address.

When a PE/ABR/ASBR/RBR receives an I/S-PMSI A-D route with the Controller EC, it MUST originate a corresponding Leaf A-D route. The PTA from the I/S-PMSI A-D route is copied to the Leaf A-D route, and an IP Address Specific Route Target to attached to the Leaf A-D route. The Global Administrator field of the RT is set to the address of the controller (as encoded in the received Controller EC), and the Local Administrator field is set to 0.

Note that, the above is done even if the Leaf Information Required (LIR) bit in the Flags field of the I/S-PMSI A-D route's PMSI Tunnel Attribute (PTA) is not set. If the LIR bit in the Flags field of the I/S-PMSI A-D route's PTA is set, then the above mentioned RTs are in addition to the RT that the PE attaches according to the procedures in [RFC6514], [RFC7524], or [I-D.ietf-bess-evpn-bum-procedure-updates]. In other words, the Leaf A-D route will have RTs for both the controllers and the upstream PE or segmentation points in this case.

When a controller receives the advertisement and/or withdrawl of Leaf A-D routes, it derives the set of leaves for the tunnel identified in the PTA, calculate and set up the tree according to procedurs in [I-D.ietf-bess-bgp-multicast-controller] or [I-D.ietf-pim-sr-p2mp-policy]. The controller does not further propagate the received advertisement and/or withdrawl, unless there are other RTs attached.

3.3. Controller Originated I/S-PMSI Routes

When I/S-PMSI A-D routes are to be originated from the controllers, it is expected that the controller, based on central planning, has the knowledge of each VPN/BD's Route Target, each PE's RD for the VPN/BD, and the Tunnel Type and Identifier for each I/S-PMSI. If the tunnel aggregation is used, the controllers also allocate labels from the DCB for the I/S-PMSIs.

The controller constructs the I/S-PMSI A-D route the same way as if an ingress PE would be originating the routes. There are some exceptions in case inter-AS/region segmentation is used, as specified in Section 3.3.1.

Specifically, the controller uses the ingress PE's RD and RTs for the VPN/BD, and use the ingress PE's address as "Originating Router's IP Address" when constructing the I/S-PMSI A-D routes. The routes are sent with the controller's address as next-hop initially, though the next-hop may change as the routes propagates.

When the Ingress PE router receives the I/S-PMSI A-D routes, it sets up corresponding forwarding state as if it originated the routes per its local provisioning. Note that the next-hop address of the routes will be different from the case where the ingress PE originates the routes, but that does not matter.

3.3.1. Inter-AS/Region Segmentation

In case of segmentation, instead of using the Route Target for the VPN/BD, the controller constructs an IP Address specific Route Target with the Global Administrator Field set to the corresponding ingress PE's address and the Local Administrator Field set to 0. This targets the I/S-PMSI A-D routes to the Ingress PEs only.

The controller also sets the Originating Router's IP Address field of the I/S-PMSI A-D route to its own address.

The receiving Ingress PE associate the I/S-PMSI A-D route to the corresponding VRF/BD based on the RD of the received route. It then re-originate a corresponding I/S-PMSI A-D route based on the

received I/S-PMSI A-D route from the controller by doing the following:

*Changing the Originating Router's IP address to its own

*Replacing the Route Target with the Route Target for the VPN/BD

3.4. Automatic DCB Label Allocation by Controllers

If it is desired for a PE to originate I/S-PMSI A-D routes on its own but with DCB labels dynamically allcated by a controller, the PE originates the I/S-PMSI A-D route with the Tunnel Type in the PTA set to "no tunnel information present", the LIR bit in the PTA'S Flags field set to 1, and attaches an IP Address Specific RT. The RT's Global Administrator Field is set to the Controller's address and Local Administrator field is set to 0.

When the controller receives the I/S-PMSI A-D route, it allocates a DCB label and responds with a Leaf A-D route. The Label field of the Leaf A-D route's PTA is set to the allocated DCB label.

When the PE receives the Leaf A-D route, it re-advertises the I/S-PMSI A-D route, with an additional RT for the corresponding VPN/BD. The PTA's tunnel information is set as needed and the Label field is set to the DCB label received in the Leaf A-D route. The LIR bit in the Flags field of the PTA is set to 1 or 0 as needed. If it is set to 0, the controller withdraws the Leaf A-D route but does not release the allocated label.

When the PE withdraws the I/S-PMSI A-D route, the controller release the DCB label and withdraws the corresponding Leaf A-D route if it had not been withdrawn before.

4. Security Considerations

This document does not change security aspects as discussed in [RFC4360], [6514], [7432], and [I-D.ietf-bess-evpn-bum-procedure-updates].

5. IANA Considerations

To be added.

6. Acknowledgements

7. References

7.1. 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", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-bum-procedure-updates-11, 7 October 2021, <https://www.ietf.org/archive/id/draft-ietf-bess-evpn-bum-procedure-updates-11.txt>.

[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>.

[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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

7.2. Informative References

[I-D.ietf-bess-bgp-multicast-controller] Zhang, Z., Raszuk, R., Pacella, D., and A. Gulko, "Controller Based BGP Multicast Signaling", Work in Progress, Internet-Draft, draft-ietf-bess-bgp-multicast-controller-07, 12 July 2021, <https://www.ietf.org/archive/id/draft-ietf-bess-bgp-multicast-controller-07.txt>.

[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", Work in Progress, Internet-Draft, draft-ietf-bess-mvpn-evpn-aggregation-label-06, 19 April 2021, <https://www.ietf.org/archive/id/draft-ietf-bess-mvpn-evpn-aggregation-label-06.txt>.

[I-D.ietf-bess-mvpn-evpn-sr-p2mp] Parekh, R., Filsfils, C., Venkateswaran, A., Bidgoli, H., Voyer, D., and Z. Zhang, "Multicast and Ethernet VPN with Segment Routing P2MP", Work in Progress, Internet-Draft, draft-ietf-bess-mvpn-evpn-sr-p2mp-04, 19 October 2021, <https://www.ietf.org/archive/id/draft-ietf-bess-mvpn-evpn-sr-p2mp-04.txt>.

[I-D.ietf-pim-sr-p2mp-policy] (editor), D. V., Filsfils, C., Parekh, R., Bidgoli, H., and Z. Zhang, "Segment Routing Point-to-Multipoint Policy", Work in Progress, Internet-Draft, draft-ietf-pim-sr-p2mp-policy-03, 23 August 2021,

<https://www.ietf.org/archive/id/draft-ietf-pim-sr-p2mp-policy-03.txt>.

[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>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks

Email: zzhang@juniper.net

Rishabh Parekh
Cisco Systems

Email: riparekh@cisco.com

Zheng Zhang
ZTE

Email: zhang.zheng@zte.com.cn

Hooman Bidgoli
Nokia

Email: hooman.bidgoli@nokia.com