Abstract

This draft proposes an extension to [RFC7432] to do ARP synchronizing and IP aliasing for Layer 3 routes that is needed for pure L3 EVPN to build a complete IP ECMP. The phrase "pure L3 EVPN" means that there is no MAC-VRF or IRB interface in the use case.

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 May 31, 2020.

Copyright Notice

Copyright (c) 2019 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 (https://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 document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.
1. Introduction

[I-D.sajassi-bess-evpn-ip-aliasing] proposes an extension to [RFC7432] to do aliasing for Layer 3 routes that is needed for symmetric IRB to build a complete IP ECMP. But typically there may be both IRB interfaces (to do EVPN IRB per-MAC-VRF basis) and VRF-interfaces in the same IP-VRF instance. It is necessary to apply the EVPN control-plane to the VRF-interfaces in order to support both such situations and the pure L3 EVPN use case where no IRB interfaces will be found in the IP-VRF instances.

![Figure 1: ARP/ND Synchronizing and IP Aliasing without IRB](image-url)
Consider a pair of multi-homed PEs PE1 and PE2. Let there be two hosts H1 and H2 attached to them via a L2 switch SW1. Consider another PE PE3 and a host H3 attached to it. The H1 and H2 represent subnet SN1 and the H3 represents subnet SN2.

Note that it is different from [I-D.sajassi-bess-evpn-ip-aliasing] in the following aspects: There is no MAC-VRF or IRB interface on PE1/PE2/PE3. And it is the IP-VRFs that are called as EVPN instance instead. Such EVPN instance can be called pure L3 EVPN instance or L3 EVI for short. The anycast gateway of H1/H2 is configured on a sub-interface on PE1/PE2.

Note that the communication between H1 and H2 won’t pass through any of the multi-homed PEs. So it is not necessary for PE1/PE2 keeping a Broadcast domain and its IRB for SN1.

Note that the SW1 multi-homing PE1 and PE2 via a LAG interface which maybe load-balance traffic to the PEs.

This draft proposes an extension to do ARP/ND synchronizing and IP aliasing for Layer 3 routes that is needed for L3 EVI to build a complete IP ECMP.

1.1. Terminology

Most of the terminology used in this documents comes from [RFC7432] and [I-D.sajassi-bess-evpn-ip-aliasing] except for the following:

VRF Interface: A interface that connects to a CE for an IP-VRF but is not an IRB interface.

L3 EVI: An EVPN instance spanning the Provider Edge (PE) devices participating in that EVPN which contains VRF Interfaces and maybe contains IRB interfaces.

EAD/IP-VRF: Ethernet Auto-Discovery route per IP-VRF, which differentiates itself from IP-EAD/EVI route by the MPLS label field.

RMAC: Router’s MAC, which is signaled in the Router’s MAC extended community.

2. ARP/ND Synching and IP Aliasing

Host IP and MAC routes are learnt by PEs on the access side via a control plane protocol like ARP. In case where a CE is multihomed to multiple PE nodes using a LAG and is running in All-Active Redundancy Mode, the Host IP will be learnt and advertised in the MAC/IP Advertisement only by the PE that receives the ARP packet. The MAC/
IP Advertisement with non-zero ESI will be received by both PE2 and PE3.

As a result, after PE2 receives the MAC/IP Advertisement and imports it to the L3 EVI, PE2 installs an ARP entry to the VRF interface whose subnet matches the IP Address from the MAC/IP Advertisement. Such ARP entry is called remote synched ARP Entry in this document.

Note that the PEs follow [I-D.sajassi-bess-evpn-ip-aliasing] to achieve the ESI load balance except for the constructing of MAC/IP Advertisement Route and IP-EAD/EVI route.

When PE3 load balance the traffic towards the multihomed Ethernet Segment, both PE1 and PE2 would have been prepared with corresponding ARP entry yet because of the ARP synching procedures.

It is important to explain that typically there may be both IRB interface and VRF interface in an IP-VRF instance, which is called as the "VRF interface in EVPN IRB" use-case in this document. But each IRB/VRF interface is independent to each other in EVPN control plane. So the use-case here is constrained to a pure L3 EVPN schema. Because it is enough to describe all the control-plane updates for both the pure L3 EVPN use-case and the "VRF interface in EVPN IRB" use-case.

In current EVPN control-plane for "VRF interface in EVPN IRB" use-case, the VRF interface is considered as "external link" and it just inter-operates with the EVPN control-plane. But in this document it is assumed to be better if the EVPN control-plane directly applied to the VRF interface.

2.1. Constructing MAC/IP Advertisement Route

This draft introduces a new usage/construction of MAC/IP Advertisement route to enable Aliasing for IP addresses in pure L3 EVPN use-cases. The usage/construction of this route remains similar to that described in RFC 7432 with a few notable exceptions as below.

* The Route-Distinguisher should be set to the corresponding L3VPN context.

* The Ethernet Tag should be set to 0.

* The MAC/IP Advertisement SHOULD carry one or more IP VRF Route-Target (RT) attributes.

* The ESI SHOULD be set to the ESI of the VRF interface from which the ARP entry is learned.
Note that the ESI is not used to install remote synched ARP entries to corresponding VRF interfaces on PE1/PE2. It is only used to load balance traffic on PE3.

* The MPLS Label1 should be set to implicit-null in MPLS/SRv6 encapsulation. For VXLAN encapsulation, the MPLS label1 should be set to 0 instead. Note that no MAC-VRF can be found here.

* The MPLS Label2 should be set to the local label of the IP-VRF in MPLS or VXLAN EVPN. But it should be set to implicit-null in SRv6 EVPN.

Note that the label may be VNI label or MPLS label.

Note that in SRv6 EVPN an SRv6 L3 Service TLV MAY also be advertised along with the route following [I-D.dawra-bess-srv6-services]. But SRv6 L2 Service TLV won’t be advertised along with the route. Because that no MAC-VRF exists in the use case.

* The RMAC Extended Community attribute SHOULD be carried in VXLAN EVPN.

2.2. Constructing EAD/IP-VRF Route

Note that the MAC/IP Advertisement is used for two reasons. It is used between PE1 and PE2 to synch the ARP entries to each other. It is used between PE1/PE2 and PE3 to achieve the load balance to ES adjacent PEs.

The usage/construction of this route is similar to the IP-EAD/EVI route described in [I-D.sajassi-bess-evpn-ip-aliasing] with a few notable exceptions as below.

* The MPLS Label should be set to the local label of the IP-VRF in MPLS EVPN or VXLAN EVPN. But it should be set to implicit-null in SRv6 EVPN.

Note that in SRv6 EVPN an L3 Service SID MAY also be advertised along with the route following [I-D.dawra-bess-srv6-services].

Such Ethernet Auto-Discovery route is called Ethernet Auto-Discvoery route per IP-VRF which is abbreviated as EAD/IP-VRF in this document.

2.3. Constructing EAD/ES Route

The usage/construction of this route remains similar to that described in section 3.1.1. of [I-D.sajassi-bess-evpn-ip-aliasing] with a few notable exceptions as explained as below.
There may be no MAC-VRF RTs in the EAD/ES Route.

3. Fast Convergence for Routed Traffic

The procedures for Fast Convergence do not change from [I-D.sajassi-bess-evpn-ip-aliasing] except for a few notable exceptions as explained as below.

The local ARP entries and remote synced ARP entries is installed/learned on a VRF interface rather than an IRB interface.

There is no MAC entry.

4. Determining Reach-ability to Unicast IP Addresses

The procedures for local/remote host learning and MAC/IP Advertisement route constructing are described above. The procedures for Route Resolution do not change from [I-D.sajassi-bess-evpn-ip-aliasing].

5. Forwarding Unicast Packets

Because of the nature of the MPLS label or SRv6 SID for IP-VRF instance, when these EAD/IP-VRF routes are referred in IP-VRF routing and forwarding procedures, the inner ethernet headers are absent on the corresponding packets transported following these EAD/IP-VRF routes.

Note that in [I-D.sajassi-bess-evpn-ip-aliasing] the inner ethernet header need to be included in the packets which are sent from IP-VRF following IP-EAD/EVI routes, because that the MPLS label of such IP-EAD/EVI route is for MAC-VRF, not for IP-VRF. and the inner destination MAC of these packets is following the "Router’s MAC" extended community of MAC/IP advertisement routes with non-zero ESI. But in many use cases, the RMAC is not the same as the IRB interface’s own MAC and the RMAC is not the same among different PEs. For example, the MAC address of the two IRB interfaces with anycast GW-IP address will be the same, but these two IRB interfaces lies on two different GW node and their "Router’s MAC" is typically not the same. In these use cases, it is recommended to use EAD/IP-VRF route instead, even if there is indeed a MAC-VRF instance.

Note that in EVPN IRB use cases, the EAD/IP-VRF route is more accordant with the symmetric IRB concept in the sense of data-plane behavior for unicast packets than the IP-EAD/EVI route of [I-D.sajassi-bess-evpn-ip-aliasing].
Note that from the viewpoint of the route receiver it is impossible to distinguish the EAD/IP-VRF route from the IP-EAD/EVI route. So the receiver has to configure how to interpret the remote EAD/EVI route. If it is interpreted as EAD/IP-VRF route, the corresponding transported unicast packets will not be inserted with an Ethernet header. But if it is interpreted as IP-EAD/EVI route, they will.

Note that we will rely on such configuration only in MPLS/SRv6 EVPN, it is not needed in VXLAN EVPN.

6. Use RT-5 Route with ESI to Reduce Route Advertisement

Given that PE1/PE2 can receive remote synced ARP entries from each other by RT-2 route following section 2.1. So it is not necessary for PE1/PE2 to advertise per-host IP prefixes by RT-2 routes. It is recommended that PE1/PE2 advertise an RT-5 route per subnet to PE3 instead. The ESI of these RT-5 routes can be set to the ESI of the corresponding VRF interface. If the VRF interface fails, these subnets will achieve more faster convergency on PE3 by the withdraw of the corresponding EAD/IP-VRF route.

7. Load Balancing of Unicast Packets

It is the same as [I-D.sajassi-bess-evpn-ip-aliasing].

8. Security Considerations

This document does not introduce any new security considerations other than already discussed in [RFC7432] and [RFC8365].

9. IANA Considerations

There is no IANA consideration.

10. Normative References

[I-D.dawra-bess-srv6-services]
Dawra, G., Filsfils, C., Brissette, P., Agrawal, S.,
Leddy, J., daniel.boyer@bell.ca, d.,
daniel.bernier@bell.ca, d., Steinberg, D., Raszuk, R.,
Decraene, B., Matsushima, S., Zhuang, S., and J. Rabadan,
"SRv6 BGP based Overlay services", draft-dawra-bess-
srv6-services-02 (work in progress), July 2019.

[I-D.ietf-bess-evpn-prefix-advertisement]
Rabadan, J., Henderickx, W., Drake, J., Lin, W., and A.
Sajassi, "IP Prefix Advertisement in EVPN", draft-ietf-
bess-evpn-prefix-advertisement-11 (work in progress), May 2018.
[I-D.sajassi-bess-evpn-ip-aliasing]


Authors’ Addresses

Yubao(Bob) Wang
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: yubao.wang2008@hotmail.com

Zheng(Sandy) Zhang
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: zzhang_ietf@hotmail.com