Workgroup: BESS Working Group
Internet-Draft:
draft-mackenzie-bess-evpn-l3mh-proto-04
Published: 7 February 2024
Intended Status: Standards Track
Expires: 10 August 2024
Authors: M. MacKenzie, Ed.  P. Brissette, Ed.
Cisco  Cisco
S. Matsushima  W. Lin  J. Rabadan
Softbank  Juniper  Nokia

# EVPN multi-homing support for L3 services

## Abstract

This document introduces the utilization of EVPN Multi-Chassis Link Aggregation Group (MC-LAG) technology to enhance network availability and load balancing for various L3 services in EVPN.

## 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 10 August 2024.

## Copyright Notice

**Table of Contents**

## 1. Introduction

Resilient L3VPN service to a CE requires multiple service PEs to run a Multi-Chassis Link Aggregation Group mechanism, which previously required a proprietary ICL control plane link between them.

This extension to [RFC9135] and to [RFC9136] brings EVPN based MC-LAG all-active multi-homing load-balancing to various services (L2 and L3) delivered by EVPN. Although this solution is also applicable to some L2 service use cases, (example Centralized Gateway) this document focuses on the L3VPN [RFC4364] use case to provide examples.

EVPN ESI-LAG is completely transparent to a CE device, and provides link and node level redundancy with load-balancing using the existing BGP control plane required by the L3 services.

For example, the L3VPN service can be MPLS, VxLAN or SRv6 based, and does not require EVPN signaling to remote neighbors. The EVPN signaling is limited to the redundant service PEs sharing a Ethernet Segment Identifier (ESI). This is used to synchronize ARP/ND, multicast Join/Leave, and IGP routes replacing need for ICL link.

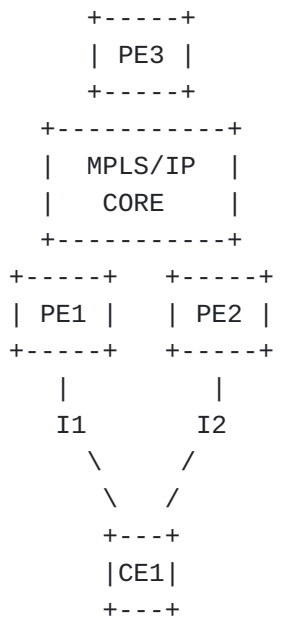



Figure 1: EVPN MC-LAG Topology

Figure 1 shows a MC-LAG multi-homing topology where PE1 and PE2 are part of the same redundancy group providing multi-homing to CE1 via interfaces I1 and I2. PE1, PE2 and PE3 are attached to the same L3VPN thru the core (running [RFC4364] and/or [RFC9136] procedures). Interfaces I1 and I2 are Bundle-Ethernet interfaces running LACP protocol. The CE device can be a layer-2 or layer-3 device

connecting to the redundant PEs over a single LACP LAG port. In the case of a layer-3 CE device, this document looks to solve the case of an IGP adjacency between PEs and CE. Further study is needed to support BGP PE to CE protocols. The core, shown as IP or MPLS enabled, provides wide range of L3 services. MC-LAG multi-homing functionality is decoupled from those services in the core and it focuses on providing multi-homing to CE.

To deliver resilient layer-3 services and provide traffic load-balancing towards the access, the two service PEs advertise layer-3 reach-ability towards the layer-3 core and both be eligible to receive traffic and forward towards the Access.

### 1.1. Problems with unicast load-balancing from core to CE

The layer-2 hashing performed by CE over its LAG port means that its possible for only one service PE to populate its ARP/ND cache. Take for example PE1 and PE2 from Figure 1. If CE1 ARP/ND response happens to always hash over I1 towards PE1, then PE2 ARP/ND table remains empty. Since unicast traffic from remote PEs can be received by either service PE, traffic that reaches the service PE2 does not find an ARP entry matching the host IP address and traffic is dropped until its ARP/ND table is updated.

If the CEs hash implementation always calculates the ARP/ND response towards PE1, the resolution on PE2 never succeeds and traffic load balanced to PE2 is permanently dropped.

The route sync solution is described in Section 2.5

### 1.2. Problems with multicast from core to CE

Like the unicast behavior above, multicast IGMP/MLD join messages from CE to LAG link may always hash to a single PE.

When PIM runs on both redundant layer-3 PEs, both serving multicast for the same access segment, PIM hello messages [RFC7761] issued by I1 (Figure 1) are not received by I2, and, vice versa; PIM hello messages issued by I2 are not received by I1. This is due to the CE not being able to switch traffic between the two members of the same LAG. Both PEs therefore become PIM Designated Router (DR). The PIM DR is responsible for tracking local multicast listeners and forwarding traffic to those listeners. The PIM DR is also responsible for sending local Join/Prune messages towards the RP or source. However, due to the CE hashing, a particular IGMP join for a given multicast group is received by only one of the PEs. Only that PE programs the multicast route for the group and issues a PIM join message.

The multicast route sync solution is described in Section 2.6

### 1.3. Problems with IGP adjacencies over the LAG port

A layer-3 CE device/router that connects to the redundant PEs may establish an IGP adjacency on the bundle port. In this case, the adjacency is formed to one of the PEs and IGP customer route(s) is only present on that PE.

This prevents the load-balancing benefits of redundant PEs from supporting this use case, as only one PE is aware and advertising the customer routes to the core.

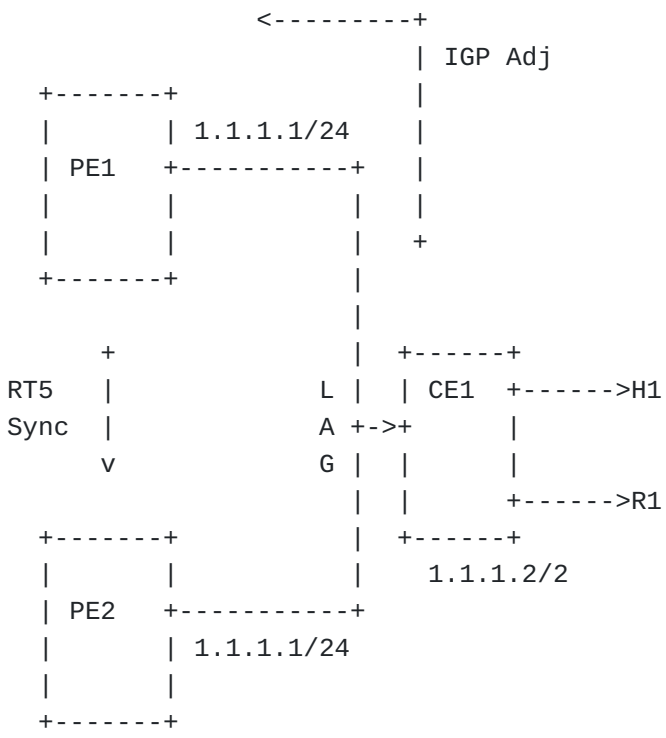



Figure 2: IGP Adjacency over LAG Port

Figure 2 provides an example of this use case, where CE1 forms an IGP adjacency with PE1 (example: ISIS or OSPF), and advertises its H1 and R1 routes into the IP-VRF of PE1. PE1 may then redistribute this IGP route into the core as an L3 service. Any remote PEs are only aware of the service from PE1, and cannot load balance through PE2 as well.

Further study is required to support the case of BGP PE to CE protocols.

A solution to this is described in Section 2.7

### 1.4. Problems with supporting multiple subnets on same ES in all active mode

In the case where the L3 service is L3VPN such as [RFC4364], it is likely the CE device could be a layer-2 switch supporting multiple subnets through the use of VLANs. In addition, each VLAN may be associated with a different customer VRF.

When ARP/ND routes are synchronized between the PEs for ARP proxy support using RT-2, a similar problem is encountered as described by Section 1.1 of [I-D.ietf-bess-evpn-ac-aware-bundling]. The PE receiving RT-2 is unable to determine which sub-interface the ARP/ND entry is associated with.

When IGMP/MLD routes are synchronized between the PEs using RT-7 and RT-8, a similar problem is encountered as described by Section 1.2 of [I-D.ietf-bess-evpn-ac-aware-bundling]. The PE receiving RT-7 and RT-8 is unable to determine which sub-interface the IGMP join is associated with.

This document proposes to use the solution defined by Section 4 of [I-D.ietf-bess-evpn-ac-aware-bundling] to solve both these cases. All route sync messages (RT-2, RT-5, RT-7, RT-8) carry an Attachment Circuit Identifier Extended Community to signal which sub-interface the routes were learnt on.

This document focuses on configuration models over access-facing interfaces with L3 sub-interfaces. Models with both L2 and L3 sub interfaces on a interface are left for future study.

### 1.5. Acronyms

**BD:** Broadcast Domain

**BE:** Bundle Ethernet

**DF:** Designated Forwarder

**DR:** Multicast Designated Router

**EC:** BGP Extended Community

**ES:** Ethernet Segment. When a customer site (device or network) is connected to one or more PEs via a set of Ethernet links, then that set of links is referred to as an 'Ethernet Segment'.

**ESI:** Ethernet Segment Identifier. A unique non-zero identifier that identifies an Ethernet Segment is called an 'Ethernet Segment Identifier'.

**ESI-LAG:**
This refers to multi-homing scenario where peering PEs, connected to same CE, are two, three or more.

**ETAG:** Ethernet Tag. An Ethernet tag identifies a particular broadcast domain, e.g., a VLAN. An EVPN instance consists of one or more broadcast domains.

**EVI:** An EVPN instance spanning the Provider Edge (PE) devices participating in that EVPN. It is used to assist a L3 VRF for route synchronization.

**GRT:** Global Routing Table

**ICL:** Inter Chassis Link

**IGMP:** Internet Group Management Protocol

**IGP:** Interior Gateway Protocol

**IP-VRF:** A VPN Routing and Forwarding table for IP routes on an PE. The IP routes could be populated by EVPN and IP-VPN address families. An IP-VRF is also an instantiation of a layer 3 VPN in an PE.

**L3AA** All-Active Redundancy Mode for Layer 3 services. When all PEs attached to an Ethernet segment are allowed to forward known unicast traffic to/from that Ethernet segment for a given VLAN,

then the Ethernet segment is defined to be operating in All-Active redundancy mode.

**MAC-VRF:** A Virtual Routing and Forwarding table for Media Access Control (MAC) addresses on a PE. A MAC-VRF is also an instantiation of an EVI in a PE

**MC-LAG:** Multi-Chassis Link Aggregation Group (MC-LAG).

**MLD:** Multicast Listener Discovery.

**PE:** Provider Edge.

**PIM:** Protocol Independent Multicast.

**RD:** Route Distinguisher used in BGP.

**RP:** Multicast Rendezvous Point.

**RT:** Route-Targets used in BGP

**RT-2:** EVPN route type 2, i.e., MAC/IP advertisement route, as defined in [RFC7432].

**RT-5:** EVPN route type 5, i.e., IP Prefix route, as defined in Section 3 of [RFC9136].

**RT-7:** EVPN route type 7, i.e., Multicast Join Synch Route, as defined in Section 9.2 of [RFC9251].

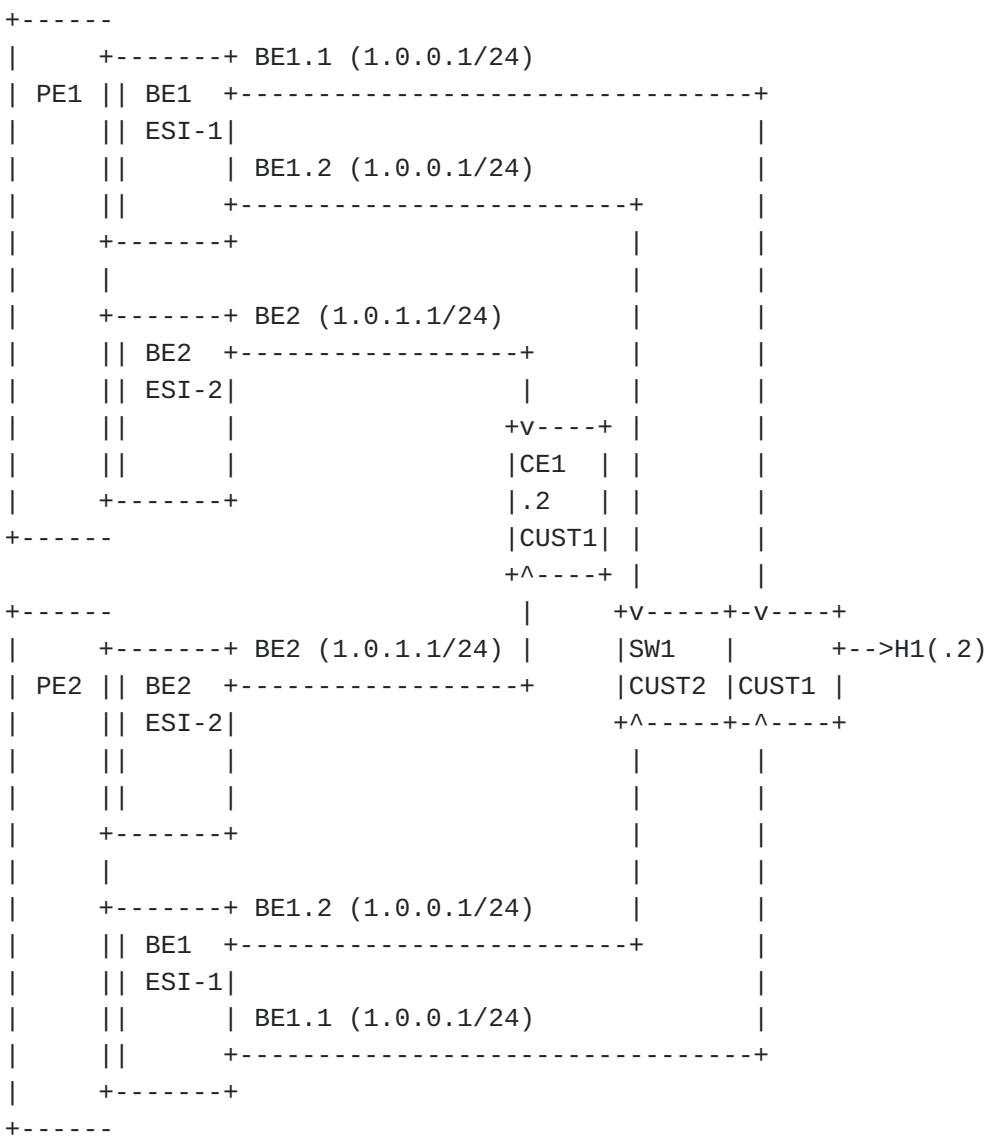
**RT-8:** EVPN route type 8, i.e., Multicast Leave Synch Route, as defined in Section 9.3 of [RFC9251].

### 1.6. Requirements

1. The multi-homing solution MUST support Layer-3 access interface
2. The multi-homing solution MUST support Layer-3 access sub-interface
3. The solution MUST support unicast and multicast VPN services
4. The solution SHOULD support IGP synchronization
5. The solution SHOULD support unicast and multicast global routing services
6. The solution MUST support all-active load-balancing mode
7. The solution MAY support single-active load-balancing mode

8. The solution MUST support port-active load-balancing mode

## 2. Solution




PE(1,2):
CUST1-VRF (IP-VRF1)
CUST2-VRF (IP-VRF2)

SW1:
CUST1-Subnet1: 1.0.0.2/24 (VLAN 1)
CUST2-Subnet1: 1.0.0.2/24 (VLAN 2)

CE1:
CUST1-Subnet2 1.0.1.2/24

Figure 3: ARP/ND synchronization over different VRF(s)

Consider the Figure 3 topology, where two AC aware bundling service interfaces are supported. On first bundling interface BE1, PE1 and PE2 share a LAG interface with switch 1 (SW1) and have two separate (but overlapping) customer 1 and customer 2 subnets. CUST1 Subnet 1 is resolving over sub-interface VLAN 1 (.1), and CUST2 Subnet 1 is resolving over sub-interface VLAN 2 (.2).

On second bundling interface BE2, both PEs share a LAG interface with Customer Edge device 1 (CE1) and only a single Customer (CUST1) subnet on native VLAN.

Main interface BE1 on PE1 and PE2 is shared by customer 1 and 2, and represented by ESI-1.

Main interface BE2 on PE1 and PE2 is only used by customer 1, and represented by ESI-2.

If we focus on CUST1, there are 2 cases visible.

Case 1: For CE1, if its ARP requests hash towards PE2, then PE1 is unaware of its presence. For PE2 to synchronize this information to PE1, in addition to CE1 IP address (1.0.1.2) and MAC address (m1), two additional unique identifiers are needed:

1. IP-VRF. CUST 1 VRF is represented by associated L3 route targets (IP-VRF RT(s))

2. Interface. BE2 Interface is represented by ESI-2

Case 2: For Host 1 (H1), if its ARP request hash towards PE2, then PE1 is unaware of its presence. For PE2 to synchronize this information to PE1, then in addition to H1 IP address (1.0.0.2) and MAC address (m2), three additional unique identifiers are required.

1. IP-VRF. CUST 1 VRF is represented by corresponding L3 route target (IP-VRF RT(s))

2. Main Interface. BE1 Interface is represented by ESI-1

3. Sub-Interface. Subnet/VLAN 1 is represented by Attachment Circuit ID 1.

### 2.1. Usage of L3VRF route target

The synchronization of information between peering PEs is done via various EVPN route types. For instance, adjacencies in ARP/ND tables are synchronized by leveraging EVPN route type-2. When dealing with Layer-3 interface, basic principles described in [RFC9136] are leverage. By default, any routes used for synchronization are advertised with IP-VRF route targets.

Alternatively, EVPN routes may be advertised with ES-import route targets along with EVI-RT EC equal to associated IP-VRF route target. This allows BGP to distribute the route(s) to only the PEs attached to the associated ESI, and also allows routes to be applied to the respective IP-VRF(s) at receiving end.

In the example Figure 3, route synchronization from CUST1 has IP-VRF1 RT(s) and CUST2 has IP-VRF2 RT(s). As an optimization, route synchronization uses ES-import RT(s). On top of that, CUST1 has EVI-RT BGP Extended Community (EC) with IP-VRF1 RT(s), and CUST2 EVI-RT BGP Extended Community (EC) has IP-VRF2 RT(s).

### 2.2. Usage of EVPN instance

[RFC7432] eases the auto-generation of BGP constructs such as route-distinguisher and route targets per MAC-VRF, based on a unique value for the Broadcast Domain that, in this document, we referred to as EVI. Similarly as in [RFC9136], the usage of EVI is not required when dealing with L3VPN multi-homing scenarios. The RD may be auto-generated locally with a unique Id and associated RT(s) may be taken from the IP-VRF

The synchronization over GRT is different. In that specific situation, an EVPN instance may be assigned to support non-VPN layer-3 services. The assignment is only serving the purpose of providing route targets as requested by [RFC7432]; where RT(s) are mandatory per EVPN route.

EVPN enhances the multi-homing layer 3 service with the following synchronization routes:

*ARP / ND

*IGMP / MLD

*IP (for customer subnets learned from IGP adjacency)

### 2.3. Mapping for L3 Interface to ESI

The ESI represents the L3 LAG interface between PE and CEs. This ESI is signaled using RT-4 with the ES-Import Route Target as described in Section 8.1.1 of [RFC7432] so that the service PE peers can discover each other's common ES.

In the example Figure 3, route-syncs from interface BE1 have IP-VRF RT(s) or ES-Import RT and EVI-RT EC with ESI 1 as an optimization.

### 2.4. Mapping for L3 Sub-Interface to Attachment Circuit ID

The Attachment Circuit ID represents the sub-interface subnet on the L3 LAG interface between PE and CEs. The AC-ID is signaled using RT-2, RT-5, RT-7 and RT-8 by attaching Attachment Circuit ID Extended community as described in Section 6.1 of [I-D.ietf-bess-evpn-ac-aware-bundling].

In the example Figure 3, route-syncs from sub-interface BE1.1 (VLAN1) have Attachment-Circuit-ID EC with ID 1

### 2.5. Route sync for ARP/ND

This document proposes solving the issue described in Section 1.1 using RT-2 IP/MAC route sync as described in Section 10 of [RFC7432] with a modification described below.

#### 2.5.1. Local adjacency (ARP/ND) learning

In EVPN or/and EVPN-IRB ([RFC7432] or/and [RFC9135]) where multi-homing is enabled through L2 access interfaces, peering PEs learn local adjacencies upon receiving ARP and/or ND messages. Using EVPN route type-2 (MAC/IP), adjacencies are synchronized between peering PE sharing common Ethernet Segments. This allows for proper layer-2 forwarding chain establishment based on configured load-balancing mode. Locally learned MAC may also be synchronized for some Layer-2 services.

Similarly with L3 interfaces, local ARP/ND learning triggers an EVPN route type-2 synchronization to any peer PE. However, there is no need for local MAC learning or synchronization since there is no layer-2 service being offer. The MAC-only RT-2 route is NOT advertised to peer PE and L2 forwarding chains should not be programmed.

Section 9.1 of [RFC7432] describes different mechanisms to learn adjacency routes locally.

ARP/ND route synchronization (refer as ARP/ND sync route in this document), uses EVPN non-zero ESI EVPN type-2 (MAC/IP) routes to exchange between peering PE all locally learned adjacencies. Few more add-ons are needed to allow proper behavior:

* An ARP/ND Sync route SHOULD carry the IP-VRF Route Target of associated VRF

* Optionally, an ARP/ND Sync route MAY carry exactly one ES-Import Route Target extended community, the one that corresponds to the ES on which the ARP or ND was received. This is in replacement of the IP-VRF RT(s) mentioned previously. Moreover, if an ES-Import

Route Target extended community is used instead of the IP-VRF Route target, the ARP/ND Sync route MUST also carry exactly one EVI-RT extended community corresponding to the associated IP-VRF on which the ARP or ND was received. See Section 9.5 of [RFC9251] for details on how to construct the EVI-RT extended community.

* In the case where PE supports AC aware bundling, it MUST also carry one Attachment Circuit ID Extended Community. The circuit ID maps the sub-interface (or subnet) where this route was received. For details on how to encode and construct this Extended Community, see section 6.1 of [I-D.ietf-bess-evpn-ac-aware-bundling].

#### 2.5.2. Remote ARP/ND learning

When consuming a remote EVPN route type-2 synchronization route:

* BGP only imports layer-3 sync route(s) based on IP-VRF Route-targets or optionally when both ES-Import and EVI-RT extended communities match those locally configured

* The main interface is derived from the ESI

* The VLAN / sub-interface is derived from the AC-ID provided in the Attachment-Circuit-ID extended community

### 2.6. Route sync for IGMP/MLD

This document proposes solving the issue described in Section 1.2 using RT-7 and RT-8 route sync as described by [RFC9251].

Local IGMP/MLD join and leave triggers a RT-7/8 route sync to peer PE.

#### 2.6.1. Local IGMP/MLD Join/Leave learning

An IGP Join or Leave triggers a RT-7/8 route sync to any peer PE.

Section 9.1 of [RFC7432] describes different mechanisms to learn adjacency routes locally.

* As per unicast, multicast routes SHOULD carry associated IP-VRF route targets.

* Optionally, an Multicast Join or Leave Sync route MAY carry exactly one ES-Import Route Target extended community, the one that corresponds to the ES on which the IGMP/MLD Join or Leave was received.

*It MAY also carry exactly one EVI-RT EC, the one that corresponds to the associated VRF on which the IGMP Join or Leave was received. See Section 9.5 of [RFC9251] for details on how to encode and construct the EVI-RT EC.

*In case where the PE supports multiple sub-interfaces within the same Ethernet Segment, the Multicast Sync routes MUST also carry one Attachment Circuit ID extended community. The circuit ID maps the sub-interface (or subnet) this route was received. For details on how to encode and construct this Extended Community, see section 6.1 of [I-D.ietf-bess-evpn-ac-aware-bundling].

#### 2.6.2. Remote IGMP/MLD Join/Leave learning

When consuming a remote multicast RT-7 or RT-8 sync route:

*A PE only imports Multicast Sync routes received with either a Route Target or an EVI-RT that matches one of the local IP-VRF(s) (assuming the ES-import Route Target matches the Route Target of one of the local Ethernet Segments).

*The layer-3 VRF is derived from the matching EVI.

*The main interface is derived from the ESI.

*The VLAN / sub-interface is derived from the AC-ID provided in the Attachment-Circuit-ID extended community.

#### 2.6.3. Upstream PIM Join/Prune

With the IGMP join/leave sync routes, both the PEs have the membership request from a multi-homed receiver. Both the PEs are DR and send a PIM join/prune message to the RP. Both the PEs are added as leaf nodes in the multicast distribution tree. Hence, both the PEs get traffic. The PE that is the DF for the multicast flow will send the traffic on the Ethernet Segment to the receiver. The NDF PE will drop the traffic.

### 2.7. Customer Subnet Route sync using Route type-5

Section 3 of [RFC9136] provides a mechanism to synchronize layer-3 customer subnets between the PEs in order to solve problem described in Section 1.3.

Using Figure 2 as example, if PE1 forms the IGP adjacency with CE, it is the only PE with knowledge of the customer subnet R1. BGP on PE1 advertises R1 to remote PEs using L3-VPN signaling, either based on [RFC4364] IP-VPN routes or [RFC9136] EVPN IP Prefix routes.

Although PE2 has the same ES connection to the CE, and could provide load balancing to remote PEs, since it has not formed an IGP adjacency with CE, it is not aware of the customer subnet R1.

This is solved by PE1 signaling R1 to PE2 using a RT-5 synchronization route. PE2 can then advertise this customer subnet R1 towards the core as if it was locally learned through IGP, and provide load-balancing from the remote PEs. There are two possible options to achieve synchronization:

1. ESI based approach.

2. IP Gateway based approach.

### 2.7.1. ESI based approach

The procedures differ depending on whether the core is running [RFC4364] IP-VPN or the [RFC9136] EVPN IP-VRF-to-IP-VRF model:

*If the core is running [RFC4364] IP-VPN, the PE receiving the R1 IGP route from the CE advertises R1 in a RT-5 with the ESI of the Ethernet Segment, and also in an IP-VPN route. Both routes carry the IP-VRF Route Target(s). The peer PE attached to the same Ethernet Segment (PE2 in Figure 2) imports both routes for R1, but treats the non-zero ESI RT-5 as if it was a local route associated to the local Ethernet Segment. Therefore the RT-5 route is selected over the IP-VPN route for R1, and PE2 advertises a new IP-VPN route for R1 so that the remote PEs in the IP-VPN network can load balance R1 traffic to both, PE1 and PE2.

*If the core is running [RFC9136] EVPN (IP-VRF-to-IP-VRF model), the PE with the IGP adjacency (PE1) advertises R1 in a RT-5 with the corresponding ESI as before, and PE2 synchronizes the route as per section 4.2 of [I-D.ietf-bess-evpn-ip-aliasing]. The advertisement of the IP A-D routes (for the ESI) from PE1 and PE2 guarantees that the remote EVPN PEs load balance the R1 traffic to both PEs attached to the Ethernet Segment (section 4 of [I-D.ietf-bess-evpn-ip-aliasing]).

### 2.7.2. IP Gateway based approach

The procedures is very similar depending on whether the core is running [RFC4364] IP-VPN or the [RFC9136] EVPN IP-VRF-to-IP-VRF model:

*If the core is running [RFC4364] IP-VPN, the PE receiving the R1 IGP route from the CE advertises R1 in a RT-5 with the IP gateway field equal to the R1 nexthop, and also a corresponding IP-VPN route. Both routes carry the IP-VRF Route Target(s). The peer PE

imports both routes for R1 where the RT-5 route is selected over the IP-VPN route for R1. Due to the adjacency synchronization done via EVPN RT-2, peer PE resolves R1 over the IP gateway pointing to the local interface. Peering PE advertises a new IP-VPN route for R1 so that the remote PEs in the IP-VPN network can load balance R1 traffic to both, PE1 and PE2.

*If the core is running [RFC9136] EVPN (IP-VRF-to-IP-VRF model), the mechanism works exactly like before without the need to select EVPN RT-5 over IP-VPN route. Furthermore, there is no need to generate IP-VPN route but only EVPN-RT5 for R1 so that the remote PEs can load balance R1 traffic to both, PE1 and PE2.

### 2.8. Mapping for VLAN to ETAG

[I-D.ietf-bess-evpn-ac-aware-bundling] proposes the use of an Attachment Circuit ID Extended Community to carry specific VLAN identification. To avoid the usage of EC, the Ethernet-tag field may be used to signal VLAN/sub-interface identification between service PE peers in RT-2, RT-5, RT-7 and RT-8 as opposed to the Attachment Circuit Extended Community.

## 3. Extensions to RT-2, RT-5, RT-7 and RT-8

This document proposes extending the use case of Extended communities already defined in other drafts for the route types RT-2, RT-5, RT-7 and RT-8.

*EVI-RT Extended Community as defined in Section 9.5 of [RFC9251].

*Attachment Circuit ID Extended Community as defined in Section 6.1 of [I-D.ietf-bess-evpn-ac-aware-bundling].

## 4. Convergence Considerations

Left for future study.

## 5. Overall Advantages

The use of EVPN ESI-LAG all active multi-homing brings the following benefits to L3 BGP services:

*Open standards based per interface all-active redundancy mechanism that eliminates the need to run ICCP and LDP.

*Agnostic of underlay technology (MPLS, VXLAN, SRv6) and associated services (L3, L3-VPN).

*Replaces legacy MC-LAG ICCP-based solution, and offers following additional benefits:

  -Fast convergence with mass-withdraw is possible with EVPN.

  -Avoid the need of a dedicated ICCP channel between peering PEs.

*Requires signaling already defined in existing EVPN RFCs [RFC7432], [RFC9136], [RFC9251] and draft [I-D.ietf-bess-evpn-ac-aware-bundling].

*Removes the burden of having the need for ICL link and any proprietary protocols.

## 6. Security Considerations

The same Security Considerations described in [RFC7432] are valid for this document.

## 7. IANA Considerations

There are no IANA considerations.

## 8. References

### 8.1. Normative References


**[I-D.ietf-bess-evpn-ac-aware-bundling]**
Sajassi, A., Mishra, M. P., Thoria, S., Rabadan, J., and J. Drake, "AC-Aware Bundling Service Interface in EVPN", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-ac-aware-bundling-04, 15 November 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-evpn-ac-aware-bundling-04>.

**[I-D.ietf-bess-evpn-ip-aliasing]**
Sajassi, A., Rabadan, J., Pasupula, S., Krattiger, L., and J. Drake, "EVPN Support for L3 Fast Convergence and Aliasing/Backup Path", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-ip-aliasing-00, 1 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-evpn-ip-aliasing-00>.

**[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>.

**[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>.

**[RFC9135]** Sajassi, A., Salam, S., Thoria, S., Drake, J., and J. Rabadan, "Integrated Routing and Bridging in Ethernet VPN (EVPN)", RFC 9135, DOI 10.17487/RFC9135, October 2021, <https://www.rfc-editor.org/info/rfc9135>.

**[RFC9136]** Rabadan, J., Ed., Henderickx, W., Drake, J., Lin, W., and A. Sajassi, "IP Prefix Advertisement in Ethernet VPN (EVPN)", RFC 9136, DOI 10.17487/RFC9136, October 2021, <https://www.rfc-editor.org/info/rfc9136>.

**[RFC9251]** Sajassi, A., Thoria, S., Mishra, M., Patel, K., Drake, J., and W. Lin, "Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN)", RFC 9251, DOI 10.17487/RFC9251, June 2022, <https://www.rfc-editor.org/info/rfc9251>.

### 8.2. Informative References

**[RFC4364]** Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

**[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, <https://www.rfc-editor.org/info/rfc7432>.

**[RFC7761]** Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March 2016, <https://www.rfc-editor.org/info/rfc7761>.

## Appendix A. Contributors

The following people has contributed substantially to this document:

Jiri Chaloupka
Cisco
Email: jichalou@cisco.com

Jayashree Subramanian
Cisco
Email: jays@cisco.com

## Authors' Addresses

Michael MacKenzie (editor)
Cisco Systems

Email: mimacken@cisco.com

Patrice Brissette (editor)
Cisco Systems

Email: pbrisset@cisco.com

Satoru Matsushima
Softbank

Email: satoru.matsushima@g.softbank.co.jp

Wen Lin
Juniper

Email: wlin@juniper.com

Jorge Rabadan
Nokia

Email: jorge.rabadan@nokia.com