

BESS Workgroup  
Internet-Draft  
Intended status: Standards Track  
Expires: 19 December 2025

R. Bickhart  
Twitch  
W. Lin  
Juniper Networks  
J. Drake  
Independent  
J. Rabadan  
Nokia  
A. Lo  
Arista Networks  
P. Brissette  
A. Sajassi  
Cisco  
17 June 2025

Proxy MAC-IP Advertisement in EVPNs  
draft-rbickhart-evpn-ip-mac-proxy-adv-04

Abstract

This document specifies procedures for EVPN PEs connected to a common multihomed site to generate proxy EVPN MAC-IP advertisements on behalf of other PEs to facilitate preservation of ARP/ND state across link or node failures.

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 19 December 2025.

Copyright Notice

Copyright (c) 2025 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

1. Conventions and Terminology . . . . .	2
2. Introduction . . . . .	3
3. MAC-IP Proxy Advertisements . . . . .	4
3.1. Interoperation with Legacy PEs . . . . .	6
3.2. Single-Active Multihoming Considerations . . . . .	6
3.3. MAC Route Attribute Considerations . . . . .	7
3.4. Symmetric IRB Considerations . . . . .	7
4. EVPN ARP/ND Extended Community . . . . .	8
5. IANA Considerations . . . . .	8
6. Operational Considerations . . . . .	9
7. Security Considerations . . . . .	9
8. Acknowledgements . . . . .	9
9. Normative References . . . . .	9
Authors' Addresses . . . . .	10

## 1. Conventions and Terminology

This document assumes familiarity with the terminology used in EVPN [RFC7432]. A few key terms used in this document are defined below:

CE: Customer edge device.

PE: Provider edge device.

MAC-IP: An IP address associated with a MAC address.

ARP: Address Resolution Protocol.

ND: Neighbor Discovery Protocol.

DF: Designated Forwarder.

R-bit: Router Flag in NA messages, as per ND for IPv6 [RFC4861].

O-bit: Override Flag in NA messages, as per ND for IPv6 [RFC4861].

## 2. Introduction

EVPN [RFC7432] allows for the distribution of MAC-IP bindings of connected hosts learned through IPv4 ARP and IPv6 neighbor discovery (ND) using EVPN MAC/IP advertisement routes. When end hosts are connected to a multihomed site running EVPN in all-active mode, depending on the learning mechanisms of the multihoming Provider Edge (PE) devices and the load balancing scheme implemented by the connected end hosts, local learning of a MAC-IP may be limited to a subset of the total number of multihoming PEs, possibly only a single PE device. In the steady state, the MAC-IP originally learned locally on one or more of the set of multihoming PEs is synchronized to all remaining PEs attached to the same multihoming site via the EVPN control plane. Once synchronized, the MAC-IP is available on each multihoming PE as well as other remote PEs not connected to the multihomed site for use in forwarding traffic or suppressing ARP or ND messaging in the EVPN.

When a multihoming PE suffers a node failure or its link to the Customer Edge (CE) device breaks down, any MAC-IP locally learned on that link by that PE will be invalidated and will be withdrawn from the EVPN control plane. If the PE was the only one that learned of any particular MAC-IP locally, that MAC-IP will entirely removed from both the EVPN control and forwarding plane until another PE can learn the MAC-IP again. Traffic forwarding or other EVPN features like ARP/ND suppression may fail during the intermediate period between the loss of the MAC-IP from the original local learning PE and later learning and distribution of the MAC-IP from a new local learning PE.

During the intermediate period between the loss of a MAC-IP entry from its original local learning PE and later re-learning and distribution of the MAC-IP from a new local learning PE, the ARP/ND suppression for this MAC-IP gets impacted. Meanwhile, an EVPN PE will broadcast traffic intended for this MAC to both its local access ports and all other EVPN PEs. This leads to a waste of bandwidth due to excessive flooding. Additionally, if the failed PE was the designated forwarder (DF) for the Ethernet Segment (ES), traffic will be blackholed until a new DF is elected and its forwarding plane is updated with the DF role.

There are two approaches to address this issue: globally or locally. While there could be other possible approaches, they are outside the scope of this document and are not discussed here.

In the global approach, every PE that learns a multihomed MAC-IP entry through the control plane must keep that entry for a predefined retention period after it is withdrawn by the originating PE. Here, we assume the originator is the only multihomed PE that initially

learned the MAC-IP entry from the data plane. For a remote PE not attached to the multihomed ES, the retention period must be long enough to allow at least one of the remaining multihomed PEs to relearn the MAC-IP entry in the data plane, and for the remote PE itself to learn the entry through the control plane from that multihomed PE. This approach is more susceptible to race conditions, as control plane learning depends on network scale and performance. Some remote PEs may not receive or process the EVPN Type-2 advertisement before the retention timer expires. As a result, they may remove the MAC-IP entry from their forwarding plane, which can lead to potential flooding and traffic blackholing. ARP/ND suppression is also affected. This scenario can occur in both egress node and egress link failure cases, even when fast reroute mechanisms are in place for egress link protection.

In the local approach, the issue is addressed by limiting the solution to only those PEs to which a multihomed ES is locally attached. This approach does not introduce new procedures on the remote PEs, instead focusing relearning efforts on peer multihomed PEs. As long as the MAC-IP is relearned in the data plane within a specified time by one of the remaining multihomed PEs, all remote PEs does not need to relearn the same MAC-IP entry from the control plane to retain their forwarding state for that MAC-IP entry.

This document specifies procedures that use the local approach to preserve MAC-IP state across the remaining PEs, bridging the gap between the initial failure and subsequent relearning of the MAC-IP on one of the remaining multihomed PEs. It thus avoids potential race conditions on remote PEs and minimizes disruption.

### 3. MAC-IP Proxy Advertisements

Preserving MAC-IP state in the EVPN until relearning and distribution of the new MAC-IP state to all PEs is completed can be accomplished by using MAC-IP proxy advertisements. When an MAC-IP for a host connected to a multihomed site is locally learned by a PE, the PE will advertise the MAC-IP via an EVPN MAC/IP route as usual. When other PEs learn that MAC-IP from the control plane upon reception of the MAC/IP route, they will install the ARP/ND state derived from the received MAC-IP for local use as usual. Additionally, if the receiving PE is locally connected to the same multihomed ethernet segment where the received MAC-IP originated and the MAC-IP was not previously locally learned and advertised, the receiving PE will inject its own EVPN MAC/IP route carrying the same MAC-IP (and with the same ESI) into the control plane and mark that injected route with a special proxy MAC-IP indication. Assuming that all PEs attached to the multihomed site support this proxy advertisement functionality, the result is that each PE attached to the site will

originate the given MAC-IP using an EVPN MAC/IP route, some of the route advertisements possibly carrying the proxy indication and at least one route advertisement not marked with the proxy indication.

A subsequent PE failure, link failure, or other event triggering the loss of all non-proxy MAC-IP state for a multihomed MAC-IP will cause the other PEs to start an aging timer for the proxy MAC-IP they had previously advertised. The aging timer should be initialized to the same age-time as the system default for ARP/ND aging, but an implementation may allow the initial age-time used for proxy a MAC-IP to be set administratively. While the aging timer is running, the multihoming PE will take no other actions and will continue using the proxy MAC-IP state for local forwarding and ARP/ND purposes and will continue to advertise the MAC-IP in the control plane with the proxy indication set. Remote PEs not connected to the multihomed site will ignore the proxy indication completely, and will be unaware of the difference between proxy and non-proxy MAC-IP advertisements. In this state, the EVPN will continue working as before the failure, with the exception of the failed link or PE being removed from the forwarding path.

In the event that one of the remaining multihoming PEs now learns the MAC-IP locally, it will restart the aging timer for the MAC-IP with the default ARP/ND age-time and remove the proxy indication from the EVPN MAC/IP route for the MAC-IP previously advertised in the control plane. When any other multihoming PE observes the removal of the proxy indication from at least one of the sources advertising the MAC-IP, that PE will stop the aging timer for the locally advertised proxy MAC-IP and continue advertising the MAC-IP with the proxy indication set as before. A PE may opt to accelerate the MAC-IP learning process by using a mechanism like send-refresh, as outlined in Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Network [RFC9161], before the aging timer for proxy MAC-IP entry expires.

If a multihoming PE does not manage to learn the MAC-IP locally before the aging timer for the proxy MAC-IP expires, that PE will withdraw the EVPN MAC/IP route for proxy MAC-IP that it had advertised previously. In this way, if all multihoming PEs fail to learn the MAC-IP locally within the age-time, the proxy MAC-IP advertisements will expire and be withdrawn from the PEs that had previously advertised them, removing the MAC-IP entirely from both the EVPN control and forwarding plane.

In the case that a non-proxy MAC-IP is withdrawn from the EVPN because the original dynamically learned ARP/ND entry ages out due to end host inactivity or shutdown rather than a PE node or link failure, PEs which advertised a proxy MAC-IP will still follow the

same procedures as above and retain their proxy MAC-IP advertisements until the age-time for the proxy MAC-IP has passed. Implementations may allow the proxy MAC-IP age-time to be administratively specified separately from the regular system ARP/ND age-time to tune how fast stale proxy MAC-IP advertisements are cleared from the EVPN. Additionally, a PE may optionally use a mechanism like send-refresh, as outlined in Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Network [RFC9161], to probe the liveness of the MAC-IP and withdraw the proxy MAC-IP from the control plane before the age-time if the PE determines that the MAC-IP is no longer active. Some implementations may alleviate such delay by verifying the presence of associated EVPN Ethernet Auto-Discovery per ES route, also known as the mass-withdraw route. With the presence of the mass-withdraw route, a PE may decide to remove the MAC-IP immediately to avoid potential traffic loss.

### 3.1. Interoperation with Legacy PEs

A heterogenous mix of new PEs supporting proxy MAC-IP advertisement and legacy PEs not supporting proxy MAC-IP advertisement is supported in the event of incremental configuration of the feature or incremental upgrades of PEs attached to the same ethernet segment. Although legacy PE devices will continue to operate with the traditional mechanisms and advertise only locally learned MAC-IP entries, they can make use of any remotely learned proxy MAC-IP advertised by other PEs supporting proxy advertisement. The proxy flag does not have any impact on the best path selection of EVPN MAC/IP Advertisement routes, as outlined in [I-D.ietf-bess-rfc7432bis].

### 3.2. Single-Active Multihoming Considerations

If the signaling of P/B flags is used along with the Ethernet A-D per EVI routes, as it is specified in [I-D.ietf-bess-rfc7432bis], and the remote PE is capable of processing P/B flags, proxy MAC-IP advertisement mechanism can be utilized in the single-active multihoming case. Otherwise, proxy MAC-IP advertisement is not applicable to ethernet segments configured for single-active multihoming because MAC advertisements are the indication of which multihoming PE is the DF for remote PEs not directly connected ethernet segment. Advertisement of a proxy MAC-IP by a non-DF multihoming PE will prevent remote PEs not directly attached to the ethernet segment from determining the correct DF.

### 3.3. MAC Route Attribute Considerations

When a PE advertises a proxy MAC-IP that was originally learned from the control plane with a MAC mobility extended community attached with a nonzero sequence number, the PE should advertise the new proxy MAC-IP with the same sequence number as originally received. When receiving a proxy MAC-IP with a higher sequence number, PE not attached to the same multihoming Ethernet segment withdraws its corresponding MAC/IP route regardless the state of proxy bit in its original advertisement.

When a PE advertises a proxy MAC-IP for an IPv6 address learned from the control plane that has the 'R' or 'O' bits set in the EVPN ND extended community, the new proxy MAC-IP should carry an EVPN ND extended community with the same 'R' and 'O' bits as originally received.

### 3.4. Symmetric IRB Considerations

When a PE advertises a proxy MAC-IP, it may check the configuration of the corresponding local IRB interface to determine whether IRB is operating in symmetric or asymmetric mode. In the case of symmetric IRB, the advertising PE may set the MPLS Label2 field of the MAC/IP advertisement route to either an MPLS label or a VNI corresponding to the MAC-IP's IP-VRF on the PE. When the MPLS Label2 field is populated with a VNI, the PE should additionally include the Router's MAC Extended Community carrying the MAC address of the PE originating the MAC-IP proxy advertisement. The proxy MAC/IP route is advertised with two Route Targets, one corresponding to the tenant's MAC-VRF and another corresponding to the tenant's IP-VRF.

As described in section 3 above, all hosts connected to an ES are advertised by at least one PE without the proxy indication set and also by any number of additional PEs with the proxy indication set. A remote PE can then import the proxy and non-proxy MAC-IP advertisements into its IP-VRF based on the Route Target for the tenant's IP VRF and use the MPLS label or VNI carried in the MPLS Label2 field of the MAC-IP advertisements to route IP traffic to hosts connected to the ES.

This approach does not utilize the multihoming aliasing mechanism which is provided by the ESI carried in the MAC/IP advertisement routes. Instead, IP route programming is based purely on normal IP multipath procedures using the routes imported to the IP-VRFs on remote PEs.

In the single-active case, a standby PE may use a different local preference for its proxy MAC/IP route so that remote PEs can prefer the MAC-IP advertised by the active PE in their forwarding plans. This is especially important when the remote PE only hosts the tenant's IP VRF but not the tenant's MAC-VRF.

#### 4. EVPN ARP/ND Extended Community

EVPN already defined EVPN ARP/ND Extended Community in Propagation of ARP/ND Flags in an Ethernet Virtual Private Network [RFC9047]. This document proposes an additional bit from the flags field of that community to signal the proxy advertisement state.

```

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
| Type=0x06      | Sub-Type=0x08 | Reserve|I|P|O|R| Reserved=0      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Reserved=0                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The following bits in the flags field in the third octet of the extended community are defined. The remaining bits must be set to zero when sending and must be ignored when receiving this community.

Bit	Name	Meaning
I,O,R		Defined in Propagation of ARP/ND Flags in an Ethernet Virtual Private Network [RFC9047].
P		Proxy MAC-IP advertisement defined in this document

Table 1: EVPN ARP/ND Extended Community Flags

#### 5. IANA Considerations

This document requests IANA the allocation of the flag position 5 in the ARP/ND Extended Community Flags registry located in the "Border Gateway Protocol (BGP) Extended Communities" registry.

Flag Position	Name	Reference
5	Proxy MAC-IP	This document



## 6. Operational Considerations

Depending on the number of multihoming PEs and MAC/IP scaling of an EVPN, proxy advertisement of MAC-IP entries by other PEs in addition to the devices initially learning MAC-IP entries locally in the data plane could cause scalability concerns for operators. Proxy advertisements would increase the total number of EVPN routes maintained in the route tables of PEs, as well as increase the time required for PEs to download all remotely learned EVPN routes. Protocol implementations should provide administrative controls for operators to limit proxy advertisement functionality to situations where the benefits are required and the scale overhead is acceptable.

## 7. Security Considerations

Proxy MAC-IP advertisement may potentially increase the total number of EVPN routes maintained in the control plane as it is specified in Section 6. Protocol implementations should provide administrative controls for operators to limit proxy advertisement functionality to situations where the benefits are required and the scale overhead is acceptable. Apart from that, this draft does not introduce any new security considerations to EVPN.

## 8. Acknowledgements

The authors would like to thank Alexander Vainshtein for his valuable comments and feedbacks.

## 9. Normative References

[I-D.ietf-bess-rfc7432bis]

Sajassi, A., Burdet, L. A., Drake, J., and J. Rabadan, "BGP MPLS-Based Ethernet VPN", Work in Progress, Internet-Draft, draft-ietf-bess-rfc7432bis-12, 18 March 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-bess-rfc7432bis-12>>.

[RFC4861]

Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <<https://www.rfc-editor.org/info/rfc4861>>.

[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>>.

- [RFC9047] Rabadan, J., Ed., Sathappan, S., Nagaraj, K., and W. Lin, "Propagation of ARP/ND Flags in an Ethernet Virtual Private Network (EVPN)", RFC 9047, DOI 10.17487/RFC9047, June 2021, <<https://www.rfc-editor.org/info/rfc9047>>.
- [RFC9161] Rabadan, J., Ed., Sathappan, S., Nagaraj, K., Hankins, G., and T. King, "Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Networks", RFC 9161, DOI 10.17487/RFC9161, January 2022, <<https://www.rfc-editor.org/info/rfc9161>>.

## Authors' Addresses

Ryan Bickhart  
Twitch  
Email: [rbickhar@twitch.tv](mailto:rbickhar@twitch.tv)

Wen Lin  
Juniper Networks  
Email: [wlin@juniper.net](mailto:wlin@juniper.net)

John Drake  
Independent  
Email: [je\\_drake@yahoo.com](mailto:je_drake@yahoo.com)

Jorge Rabadan  
Nokia  
Email: [jorge.rabadan@nokia.com](mailto:jorge.rabadan@nokia.com)

Alton Lo  
Arista Networks  
Email: [altonlo@arista.com](mailto:altonlo@arista.com)

Patrice Brissette  
Cisco  
Email: [pbrisset@cisco.com](mailto:pbrisset@cisco.com)

Ali Sajassi  
Cisco  
Email: [sajassi@cisco.com](mailto:sajassi@cisco.com)