bess
Internet-Draft
Intended status: Standards Track
Expires: 23 April 2026

Z. Zhang
HPE
C. Lin
New H3C Technologies
J. Rabadan
Nokia

20 October 2025

EVPN Multi-Active Multihoming
draft-zzhang-bess-evpn-multi-active-multihoming-00

Abstract

EVPN supports All-Active and Single-Active multihoming, where either all multihoming PEs are active or only one is active. In some situations, it is desired to support Multi-Active with multiple yet not all active PEs. This document specifies the Multi-Active multihoming - some multihoming PEs are in All-Active mode while others are in Single-Active mode on the same multihoming Ethernet Segment, and ingress PEs load-balance to those in All-Active mode.

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."

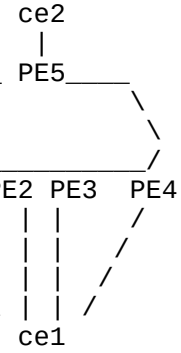
This Internet-Draft will expire on 23 April 2026.

Copyright Notice

Table of Contents

1. Introduction

Consider the following EVPN [RFC7432] multihoming situation:



CE1 is multihomed to PE1..4. It is desired that traffic from remote PEs, e.g., PE5, is load-balanced to PE1/PE2 normally, and only switch to PE3/PE4 when both PE1/PE2 are down.

This cannot be achieved via the existing All-Active or Single-Active multihoming schemes, but it can be achieved via a new Multi-Active scheme introduced in this document:

* The preferred PEs (e.g., PE1/PE2) in an MHES are considered in the All-Active mode. They set the P-bit in the L2-Attribute Extended Community to 1 and the B-bit to 0 in the Ethernet A-D per EVI routes.

* Others are considered in the Single-Active mode. They set the P-bit to 0. One of them might be the BDF (if there is only one preferred PE, which is the DF), and its B-bit is set to 1 in that case, but that has no impact with respect to the Multi-Active behavior.

* The preferred PEs are those advertising the highest (or lowest) preference value in the DF Election Extended Community.

* Remote PEs load-balance to MHES PEs that advertise P-bit 1, with backup paths to MHES PEs that advertise P-bit 0.

Note that the set of preferred PEs (and hence the rest of PEs) can change dynamically. In the above example, when PE1/PE2 goes down, PE3/PE4 becomes the preferred and advertises P-bit 1.

Whether a PE is preferred depends on its DF Preference value in the DF Election Extended Community [RFC8584]. Notice that, even if the DF Election algorithm is not Highest- or Lowest-Preference [RFC9785], the DF Preference field is still used to signal the preference for the purpose of Multi-Active multihoming.

1.1. Relationship with Single-Active, All-Active, and Port-Active

[RFC9786] specifies a Port-Active redundancy mode for multihoming, which is just the Single-Active mode without Service Carving.

Per [RFC7432]:

"The default procedure for DF election at the granularity of <ES, VLAN> for VLAN-based service or <ES, VLAN bundle> for VLAN-(aware) bundle service is referred to as "service carving". With service carving, it is possible to elect multiple DFs per Ethernet segment (one per VLAN or VLAN bundle) in order to perform load balancing of multi-destination traffic destined to a given segment. The load-balancing procedures carve up the VLAN space per ES among the PE nodes evenly, in such a way that every PE is the DF for a disjoint set of VLANs or VLAN bundles for that ES."

Per [RFC9786], Port-Active is still the Single-Active mode, but with the DF election performed per ES. As a result, the non-DFs (backup PEs) could transition the port to standby/down state because the port will not be used.

Multi-active is a mix of All-Active and Single-Active, and service carving is still applicable. Some PEs on an MHES are All-Active while others are Single-Active, and they can change dynamically.

2. Specification

For an MHES to be Multi-Active, PEs on the ES MUST be configured with appropriate preferences that are advertised in the DF Election Extended Community, even if the DF election algorithm is not Highest- or Lowest-Preference. If the DF election algorithm is Highest- or Lowest-Preference, the same preference is used for both DF election and for determining if a PE is preferred for the Multi-Active purpose.

A PE is considered preferred if one of the following conditions is met:

* When the DF election algorithm is Highest- or Lowest-Preference, the PE has the highest or lowest preference of all those advertised in the DF Election Extended Communities.

* When the DF election algorithm is neither Highest- nor Lowest-Preference, the PE has the highest preference of all those advertised in the DF Election Extended Communities.

A Multi-Active MHES may be in strict or loose mode. The above conditions are for strict mode, in which only and all the PEs with strictly highest or lowest preference are preferred. In other words, as long as one of the previously multiple preferred PEs remains, other PEs will not become preferred.

In the loose mode, there could be M preferred PEs in the MHES. Their preference values are higher (or lower) than those of others (with the IP address being the tie-breaker), but their preference values are not necessarily equal. For example, for PE1/2/3/4 with preference 100/100/80/80 in M=2 loose mode, PE1/2 are initially preferred. When PE1 goes down, PE2/3 are preferred while PE4 is not preferred (assuming PE3 has a higher IP address than PE4). For comparsion, in the strict mode, only PE2 remains preferred. If PE2 also goes down, then PE3/4 both become preferred.

The provisioning of strict/loose mode and the M value SHOULD be consistent across the PEs on an MHES. However, the information is not signaled among the PEs. In the case of inconsistency, a PE may incorrectly set/clear the P-bit, leading to the undesired load-balancing. However, that should not lead to traffic blackholing.

A preferred PE MUST operate in the All-Active mode. It sets the Multihoming Redundancy Mode bit in the ESI Label extended community to 0 (indicating All-Active), sets the P-bit in the L2-Attribute Extended Community to 1, and sets the B-bit to 0.

Otherwise, the PE MUST operate in the Single-Active mode locally, though it sets the Multihoming Redundancy Mode bit in the ESI Label extended community to 0 (indicating All-Active). It sets the P-bit in the L2-Attribute Extended Community to 0. If it is the BDF, the B-bit is set to 1.

When an ingress PE performs aliasing and load-balancing procedures for an MHES, load-balancing SHOULD be applied to all PEs setting the P-bit to 1 in the L2-Attribute Extended Community. Backup paths via PEs setting the P-bit to 0 SHOULD be set up.

3. Security Considerations

This does not introduce additional security concerns beyond what are documented in [RFC7432].

4. Acknowledgments

The authors thank Soumyodeep Joarder, Vikram Nagarajan, and Vinod Kumar Nagaraj for the discussions on the use case and solution options.

5. References

5.1. Normative References

[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>.

[RFC8584] Rabadan, J., Ed., Mohanty, S., Ed., Sajassi, A., Drake, J., Nagaraj, K., and S. Sathappan, "Framework for Ethernet VPN Designated Forwarder Election Extensibility", RFC 8584, DOI 10.17487/RFC8584, April 2019, <https://www.rfc-editor.org/info/rfc8584>.

[RFC9785] Rabadan, J., Ed., Sathappan, S., Lin, W., Drake, J., and A. Sajassi, "Preference-Based EVPN Designated Forwarder (DF) Election", RFC 9785, DOI 10.17487/RFC9785, June 2025, <https://www.rfc-editor.org/info/rfc9785>.

5.2. Informative References

[RFC9786] Brissette, P., Burdet, LA., Ed., Wen, B., Leyton, E., and J. Rabadan, "EVPN Port-Active Redundancy Mode", RFC 9786, DOI 10.17487/RFC9786, June 2025, <https://www.rfc-editor.org/info/rfc9786>.

Authors' Addresses

Zhaohui Zhang
HPE
Email: zzhang@juniper.net

Changwang Lin
New H3C Technologies
Email: linchangwang.04414@h3c.com

Jorge Rabadan
Nokia
Email: jorge.rabadan@nokia.com