Stream: Internet Engineering Task Force (IETF)
RFC: 9786
Category: Standards Track
Published: June 2025
ISSN: 2070-1721
Authors: P. Brissette *Cisco Systems*; LA. Burdet, Ed. *Cisco Systems*; B. Wen *Comcast*; E. Leyton *Verizon Wireless*; J. Rabadan *Nokia*

RFC 9786
EVPN Port-Active Redundancy Mode

Abstract

The Multi-Chassis Link Aggregation (MC-LAG) group technology enables establishing a logical link-aggregation connection with a redundant group of independent nodes. The objective of MC-LAG is to enhance both network availability and bandwidth utilization through various modes of traffic load balancing. RFC 7432 defines an EVPN-based MC-LAG with Single-Active and All-Active multihoming redundancy modes. This document builds on the existing redundancy mechanisms supported by EVPN and introduces a new active/standby redundancy mode, called 'Port-Active'.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9786.

Copyright Notice

Table of Contents

1. Introduction

EVPN [RFC7432] defines the All-Active and Single-Active redundancy modes. All-Active redundancy provides per-flow load balancing for multihoming, while Single-Active redundancy ensures service carving where only one of the Provider Edge (PE) devices in a redundancy relationship is active per service.

Although these two multihoming scenarios are widely utilized in data center and service provider access networks, there are cases where active/standby multihoming at the interface level is beneficial and necessary. The primary consideration for this new mode of load balancing is the determinism of traffic forwarding through a specific interface rather than statistical per-flow load balancing across multiple PEs providing multihoming. This determinism is essential for certain QoS features to function correctly. Additionally, this mode ensures fast convergence during failure and recovery, which is expected by customers.

This document defines the Port-Active redundancy mode as a new type of multihoming in EVPN and details how this mode operates and is supported via EVPN.

1.1. 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.

1.2. Multi-Chassis Link Aggregation (MC-LAG)

When a Customer Equipment (CE) device is multihomed to a set of PE nodes using the Link Aggregation Control Protocol (LACP) [IEEE_802.1AX_2014], the PEs must function as a single LACP entity for the Ethernet links to form and operate as a Link Aggregation Group (LAG). To achieve this, the PEs connected to the same multihomed CE must synchronize LACP configuration and operational data among them. Historically, the Inter-Chassis Communication Protocol (ICCP) [RFC7275] has been used for this synchronization. EVPN, as described in [RFC7432], covers the scenario where a CE is multihomed to multiple PE nodes, using a LAG to simplify the procedure significantly. However, this simplification comes with certain assumptions:

- A CE device connected to EVPN multihoming PEs **MUST** have a single LAG with all its links connected to the EVPN multihoming PEs in a redundancy group.
- Identical LACP parameters **MUST** be configured on peering PEs, including the system ID, port priority, and port key.

This document presumes proper LAG operation as specified in [RFC7432]. Issues resulting from deviations in the aforementioned assumptions, LAG misconfiguration, and miswiring detection across peering PEs are considered outside the scope of this document.

```
                 +-----+
                 | PE3 |
                 +-----+
              +-----------+
              |  MPLS/IP  |
              |  CORE     |
              +-----------+
            +-----+   +-----+
            | PE1 |   | PE2 |
            +-----+   +-----+
               I1       I2
                \       /
                 \     /
                  \   /
                  +---+
                  |CE1|
                  +---+
```

Figure 1: MC-LAG Topology

Figure 1 shows an MC-LAG multihoming topology where PE1 and PE2 are part of the same redundancy group providing multihoming to CE1 via interfaces I1 and I2. Interfaces I1 and I2 are members of a LAG running LACP. The core, shown as IP or MPLS enabled, provides a wide range of L2 and L3 services. MC-LAG multihoming functionality is decoupled from those services in the core, and it focuses on providing multihoming to the CE. In Port-Active redundancy mode, only one of the two interfaces, I1 or I2, would be in forwarding, and the other interface would be in standby. This also implies that all services on the active interface operate in active mode and all services on the standby interface operate in standby mode.

2. Port-Active Redundancy Mode

2.1. Overall Advantages

The use of Port-Active redundancy in EVPN networks provides the following benefits:

a. It offers open-standards-based active/standby redundancy at the interface level rather than VLAN granularity [RFC7432].
b. It eliminates the need for ICCP and LDP [RFC5036] (e.g., Virtual eXtensible Local Area Network (VXLAN) [RFC7348] or Segment Routing over IPv6 (SRv6) [RFC8402] may be used in the network).
c. This mode is agnostic of the underlying technology (MPLS, VXLAN, and SRv6) and associated services (Layer 2 (L2), Layer 3 (L3), Bridging, E-LINE, etc.)
d. It enables deterministic QoS over MC-LAG attachment circuits.
e. It is fully compliant with [RFC7432] and does not require any new protocol enhancements to existing EVPN RFCs.
f. It can leverage various Designated Forwarder (DF) election algorithms, such as modulo [RFC7432], Highest Random Weight (HRW) [RFC8584], etc.

g. It replaces legacy MC-LAG ICCP-based solutions and offers the following additional benefits:

   - Efficient support for 1+N redundancy mode (with EVPN using BGP Route Reflector), whereas ICCP requires a full mesh of LDP sessions among PEs in the redundancy group.
   - Fast convergence with mass withdraw is possible with EVPN, which has no equivalent in ICCP.

2.2. Port-Active Redundancy Procedures

The following steps outline the proposed procedure for supporting Port-Active redundancy mode with EVPN LAG:

a. The Ethernet Segment Identifier (ESI) **MUST** be assigned per access interface as described in [RFC7432]. The ESI can be auto-derived or manually assigned, and the access interface **MAY** be an L2 or L3 interface.

b. The Ethernet Segment (ES) **MUST** be configured in Port-Active redundancy mode on peering PEs for the specified access interface.

c. When ESI is configured on an L3 interface, the ES route (Route Type-4) can be the only route exchanged by PEs in the redundancy group.

d. PEs in the redundancy group leverage the DF election defined in [RFC8584] to determine which PE keeps the port in active mode and which PE(s) keeps it in standby mode. Although the DF election defined in [RFC8584] is per [ES, Ethernet Tag] granularity, the DF election is performed per [ES] in Port-Active redundancy mode. The details of this algorithm are described in Section 3.

e. The DF router **MUST** keep the corresponding access interface in an up and forwarding active state for that ES.

f. Non-DF routers **SHOULD** implement a bidirectional blocking scheme for all traffic comparable to the Single-Active redundancy mode described in [RFC7432], albeit across all VLANs.

   - Non-DF routers **MAY** bring and keep the peering access interface attached to them in an operational down state.
   - If the interface is running the LACP protocol, the non-DF PE **MAY** set the LACP state to Out of Sync (OOS) instead of setting the interface to a down state. This approach allows for better convergence during the transition from standby to active mode.

g. The primary/backup bits of the EVPN Layer 2 Attributes (L2-Attr) Extended Community [RFC8214] **SHOULD** be used to achieve better convergence, as described in Section 4.1.

3. Designated Forwarder Algorithm to Elect per Port-Active PE

The ES routes operating in Port-Active redundancy mode are advertised with the new Port Mode Load-Balancing capability bit in the DF Election Extended Community as defined in [RFC8584]. Additionally, the ES associated with the port utilizes the existing Single-Active procedure and signals the Single-Active multihomed site redundancy mode along with the Ethernet A-D per-ES route (refer to Section 7.5 of [RFC7432]). Finally, The ESI label-based split-horizon procedures specified in Section 8.3 of [RFC7432] **SHOULD** be employed to prevent transient echo packets when L2 circuits are involved.

Various algorithms for DF election are detailed in Sections 3.2 to 3.5 for comprehensive understanding, although the choice of algorithm in this solution does not significantly impact complexity or performance compared to other redundancy modes.

3.1. Capability Flag

[RFC8584] defines a DF Election Extended Community and a bitmap (2 octets) field to encode "DF Election Capabilities" to use with the DF election algorithm in the DF algorithm field:

Bit 0: D bit or 'Don't Preempt' bit, as described in [RFC9785].

Bit 1: AC-Influenced DF (AC-DF) election, as described in [RFC8584].

Bit 3: Time Synchronization, as described in [RFC9722].

```
                      1 1 1 1 1 1
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |D|A| |T| |P|                   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Amended DF Election Capabilities in the DF Election Extended Community

This document defines the following value and extends the DF Election Capabilities bitmap field:

Bit 5: Port Mode Designated Forwarder Election. This bit determines that the DF election algorithm **SHOULD** be modified to consider the port ES only and not the Ethernet Tags.

3.2. Modulo-Based Algorithm

The default DF election algorithm, or modulo-based algorithm, as described in [RFC7432] and updated by [RFC8584], is applied here at the granularity of ES only. Given that the ES-Import Route Target extended community may be auto-derived and directly inherits its auto-derived value from ESI bytes 1-6, many operators differentiate ESIs primarily within these bytes. Consequently, bytes 3-6 are utilized to determine the designated forwarder using the modulo-based DF assignment, achieving good entropy during modulo calculation across ESIs.

Assuming a redundancy group of N PE nodes, the PE with ordinal i is designated as the DF for an <ES> when (Es mod N) = i, where Es represents bytes 3-6 of that ESI.

3.3. Highest Random Weight Algorithm

An application of Highest Random Weight (HRW) to EVPN DF election is defined in [RFC8584], and it **MAY** be used and signaled. For Port-Active, this is modified to operate at the granularity of <ES> rather than per <ES, VLAN>.

Section 3.2 of [RFC8584] describes computing a 32-bit Cyclic Redundancy Check (CRC) over the concatenation of Ethernet Tag (V) and ESI (Es). For Port-Active redundancy mode, the Ethernet Tag is omitted from the CRC computation and all references to (V, Es) are replaced by (Es).

The algorithm used to determine the DF and Backup Designated Forwarder (BDF) per Section 3.2 of [RFC8584] is repeated and summarized below using only (Es) in the computation:

1. DF(Es) = Si| Weight(Es, Si) >= Weight(Es, Sj), for all j. In the case of a tie, choose the PE whose IP address is numerically the least. Note that 0 <= i,j < number of PEs in the redundancy group.
2. BDF(Es) = Sk| Weight(Es, Si) >= Weight(Es, Sk), and Weight(Es, Sk) >= Weight(Es, Sj). In the case of a tie, choose the PE whose IP address is numerically the least.

Where:

- DF(Es) is defined to be the address Si (index i) for which Weight(Es, Si) is the highest; 0 <= i < N-1.
- BDF(Es) is defined as that PE with address Sk for which the computed Weight is the next highest after the Weight of the DF. j is the running index from 0 to N-1; i and k are selected values.

3.4. Preference-Based DF Election

When the new capability 'Port Mode' is signaled, the preference-based DF election algorithm [RFC9785] is modified to consider the port only and not any associated Ethernet Tags. The Port Mode capability is compatible with the 'Don't Preempt' bit and both may be signaled. When an interface recovers, a peering PE signaling the D bit enables non-revertive behavior at the port level.

3.5. AC-Influenced DF Election

The AC-DF bit defined in [RFC8584] **MUST** be set to 0 when advertising Port Mode Designated Forwarder Election capability (P=1). When an AC (sub-interface) goes down, any resulting Ethernet A-D per EVI withdrawal does not influence the DF election.

Upon receiving the AC-DF bit set (A=1) from a remote PE, it **MUST** be ignored when performing Port Mode Designated Forwarder Election.

4. Convergence Considerations

To enhance convergence during failure and recovery when the Port-Active redundancy mode is employed, prior synchronization between peering PEs may be beneficial.

The Port-Active mode poses a challenge to synchronization since the "standby" port may be in a down state. Transitioning a "standby" port to an up state and stabilizing the network requires time. For Integrated Routing and Bridging (IRB) and L3 services, prior synchronization of ARP / Neighbor Discovery (ND) caches is recommended. Additionally, associated Virtual Routing and Forwarding (VRF) tables may need to be synchronized. For L2 services, synchronization of MAC tables may be considered.

Moreover, for members of a LAG running LACP, the ability to set the "standby" port to an "out-of-sync" state, also known as "warm-standby," can be utilized to improve convergence times.

4.1. Primary/Backup Bits per Ethernet Segment

The EVPN L2-Attr Extended Community defined in [RFC8214] **SHOULD** be advertised in the Ethernet A-D per ES route to enable fast convergence.

Only the P and B bits of the Control Flags field in the L2-Attr Extended Community are relevant to this document, specifically in the context of Ethernet A-D per ES routes:

- When advertised, the L2-Attr Extended Community **SHALL** have only the P or B bits set in the Control Flags field, and all other bits and fields **MUST** be zero.
- A remote PE receiving the optional L2-Attr Extended Community in Ethernet A-D per ES routes **SHALL** consider only the P and B bits and ignore other values.

For the L2-Attr Extended Community sent and received in Ethernet A-D per EVI routes used in [RFC8214], [RFC7432], and [RFC9744]:

- P and B bits received **SHOULD** be considered overridden by "parent" bits when advertised in the Ethernet A-D per ES.
- Other fields and bits of the extended community are used according to the procedures outlined in the referenced documents.

By adhering to these procedures, the network ensures proper handling of the L2-Attr Extended Community to maintain robust and efficient convergence across Ethernet Segments.

4.2. Backward Compatibility

Implementations that comply with [RFC7432] or [RFC8214] only (i.e., implementations that predate this specification) and that receive an L2-Attr Extended Community in Ethernet A-D per ES routes will ignore it and continue to use the default path resolution algorithms of the two specifications above:

- The L2-Attr Extended Community in Ethernet A-D per ES route is ignored.
- The remote ESI Label Extended Community [RFC7432] signals the Single-Active redundancy mode (Section 3).
- The remote Media Access Control (MAC) and/or Ethernet A-D per EVI routes are unchanged; the P and B bits in the L2-Attr Extended Community in Ethernet A-D per EVI routes are used.

5. Applicability

A prevalent deployment scenario involves providing L2 or L3 services on PE devices that offer multihoming capabilities. The services may include any L2 EVPN solutions such as EVPN Virtual Private Wire Service (VPWS) or standard EVPN as defined in [RFC7432]. Additionally, L3 services may be provided within a VPN context, as specified in [RFC4364], or within a global routing context. When a PE provides first-hop routing, EVPN IRB may also be deployed on the PEs. The mechanism outlined in this document applies to PEs providing L2 and/or L3 services where active/standby redundancy at the interface level is required.

An alternative solution to the one described in this document is MC-LAG with ICCP active/standby redundancy, as detailed in [RFC7275]. However, ICCP requires LDP to be enabled as a transport for ICCP messages. There are numerous scenarios where LDP is not necessary, such as deployments utilizing VXLAN or SRv6. The solution using EVPN, as described in this document, does not mandate the use of LDP or ICCP and remains independent of the underlay encapsulation.

6. IANA Considerations

Per this document, IANA has added the following entry to the "DF Election Capabilities" registry under the "Border Gateway Protocol (BGP) Extended Communities" registry group:

Bit	Name	Reference
5	Port Mode Designated Forwarder Election	RFC 9786

Table 1

7. Security Considerations

The security considerations described in [RFC7432] and [RFC8584] are applicable to this document.

Introducing a new capability necessitates unanimity among PEs. Without consensus on the new DF election procedures and Port Mode, the DF election algorithm defaults to the procedures outlined in [RFC8584] and [RFC7432].This fallback behavior could be exploited by an attacker who modifies the configuration of one PE within the ES. Such manipulation could force all PEs in the ES to revert to the default DF election algorithm and capabilities. In this scenario, the PEs may be subject to unfair load balancing, service disruption, and potential issues such as traffic loss or duplicate traffic, as mentioned in the security sections of those documents.

8. References

8.1. Normative References

[IEEE_802.1AX_2014] IEEE, "IEEE Standard for Local and metropolitan area networks -- Link Aggregation", IEEE 802-1ax-2014, DOI 10.1109/IEEESTD.2014.7055197, 5 March 2015, <https://ieeexplore.ieee.org/document/7055197>.

[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>.

[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>.

[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>.

[RFC8214] Boutros, S., Sajassi, A., Salam, S., Drake, J., and J. Rabadan, "Virtual Private Wire Service Support in Ethernet VPN", RFC 8214, DOI 10.17487/RFC8214, August 2017, <https://www.rfc-editor.org/info/rfc8214>.

[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>.

[RFC9722] Brissette, P., Sajassi, A., Burdet, LA., Ed., Drake, J., and J. Rabadan, "Fast Recovery for EVPN Designated Forwarder Election", RFC 9722, DOI 10.17487/RFC9722, May 2025, <https://www.rfc-editor.org/info/rfc9722>.

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

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>.

[RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed., "LDP Specification", RFC 5036, DOI 10.17487/RFC5036, October 2007, <https://www.rfc-editor.org/info/rfc5036>.

[RFC7275] Martini, L., Salam, S., Sajassi, A., Bocci, M., Matsushima, S., and T. Nadeau, "Inter-Chassis Communication Protocol for Layer 2 Virtual Private Network (L2VPN) Provider Edge (PE) Redundancy", RFC 7275, DOI 10.17487/RFC7275, June 2014, <https://www.rfc-editor.org/info/rfc7275>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC9744] Sajassi, A., Ed., Brissette, P., Uttaro, J., Drake, J., Boutros, S., and J. Rabadan, "EVPN Virtual Private Wire Service (VPWS) Flexible Cross-Connect (FXC) Service", RFC 9744, DOI 10.17487/RFC9744, March 2025, <https://www.rfc-editor.org/info/rfc9744>.

Acknowledgements

The authors thank Anoop Ghanwani for his comments and suggestions and Stephane Litkowski and Gunter Van de Velde for their careful reviews.

Contributors

In addition to the authors listed on the front page, the following people have also contributed to this document:

Ali Sajassi
Cisco Systems
United States of America
Email: sajassi@cisco.com

Samir Thoria
Cisco Systems
United States of America
Email: sthoria@cisco.com

Authors' Addresses

Patrice Brissette
Cisco Systems
Ottawa ON
Canada
Email: pbrisset@cisco.com

Luc André Burdet (EDITOR)
Cisco Systems
Canada
Email: lburdet@cisco.com

Bin Wen
Comcast
United States of America
Email: Bin_Wen@comcast.com

Edward Leyton
Verizon Wireless
United States of America
Email: edward.leyton@verizonwireless.com

Jorge Rabadan
Nokia
United States of America
Email: jorge.rabadan@nokia.com