BESS Working Group
Internet-Draft
Intended status: Informational
Expires: 25 December 2025

A. Sajassi
C. Wang
N. Malhotra
Cisco

23 June 2025

EVPN Underlay Network Migration From IPv4 to IPv6
draft-sajassi-bess-evpn-fabric-migration-02

Abstract

EVPN [RFC7432] and [RFC8365] has become the standard defacto technology/solution used in Enterprise, Data Center, and Service Provider networks because of its multi-tenancy, flexible multi-homing capabilities, efficient utilization of network bandwidth, support of workload mobility, flexible workload placement, etc. across many different types of services/VPNs. Some operators have deployed IPv4 underlay network/fabric and now plan to migrate to IPv6. This document describes the procedures to achieve such migration seamlessly.

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 25 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. Introduction

EVPN [RFC7432] and [RFC8365] has become the standard defacto technology/solution used in Enterprise, Data Center, and Service Provider networks because of its multi-tenancy, flexible multi-homing capabilities, efficient utilization of network bandwidth, support of workload mobility, flexible workload placement, etc. across many different types of services/VPNs. Some operators have deployed IPv4 underlay network/fabric and now plan to migrate to IPv6. This document describes the procedures to achieve such migration seamlessly.

It should be noted that the migration from IPv4 to IPv6 for underlay network/fabric is completely independent from overlay networks. Currently, EVPN supports natively both IPv4 and IPv6 overlay networks

(i.e., providing connectivity among IPv4 and IPv6 workloads and applications). The underlay connectivity among PE devices (aka NVEs) are established by a set of tunnels (e.g., MP2P tunnels) where the tunnel endpoints are located at the NVEs and are referred to as VTEPs (Virtual Tunnel Endpoints). In EVPN routes, VTEP address is and IP address conveyed in the Next Hop Field of the MP_REACH_NLRI attribute and it is of the same type as the underelay network (i.e., IPv4 or IPv6).

In data plane, for VxLAN encapsulation, the outer IP Header of a packet carries source and destination VTEP addresses corresponding to the underlay tunnel of type IPv4 or IPv6.

As defined in [RFC7432] and [RFC8365], the Next Hop field of the MP_REACH_NLRI attribute of the EVPN routes can be set to the IPv4 or IPv6 address of the NVE’s VTEP IP address, which allows EVPN VXLAN over IPv4 (VXLANv4) or over IPv6 (VXLANv6) Single Stack transport for greenfield deployments.

Since the EVPN VXLANv4 has been widely deployed and there is an industry trend of migrating existing IPv4 based network to IPv6, an underlay migration path from EVPN VXLANv4 to VXLANv6 for brownfield deployments must be considered. To support the fabric (i.e., underlay network) migration seamlessly and gradually (i.e., one NVE at the time), EVPN must be able to support dual IPv4/IPv6 underlay transport by carrying dual next hops so that the receiving side can choose which transport to use, and it must be backward compatible with existing single stack underlay VTEP nodes in the fabric for the purpose of gradual migration.

Section 2 discusses the procedures for seamless migration of EVPN fabric from IPv4 to Ipv6. Section 4 and Section 5 discuss how to revert the migration of EVPN fabric from IPv6 to IPv4 for a) when the migration is not complete, and b) when the migration has been completed.

1.1. Terminology

EVPN: Ethernet VPN

NVE: Network Virtualization Edge

PE: Provider Edge Device

VTEP: VXLAN Tunnel End Point

VXLANv4: EVPN VXLAN Overlay over IPv4 Underlay

VXLANv6: EVPN VXLAN Overlay over IPv6 Underlay

VXLANv4v6: EVPN VXLAN Overlay over IPv4 and IPv6 (Dual-Stack) Underlay

2. EVPN Fabric Migration from V4 to V6 for VxLAN Encapsulation

An EVPN Fabric might have hundreds or thousands of VTEPs and the fabric migration could be a long process involving software upgrade and per-VTEP migration. Since a VXLANv6 VTEP cannot communicate directly with a VXLANv4 VTEP, Single Stack Underlay will not work for Seamless (in-service) Migration, and Dual-Stack Underlay is needed for the transition.

To support In-Fabric Per VTEP Seamless Migration, a VTEP to be migrated must be upgraded to have Dual-Stack Underlay capability while other VTEPs can still run in the single stack mode without the dual-stack capability.

The following steps in Figure 1 describe in detail the procedure to migrate a VXLANv4 Fabric to a VXLANv6 Fabric:

1. Initially, all the VTEPs are running in the single-stack VXLANv4 mode. The VXLAN encapsulation among all the VTEPs is set to "ipv4’.

2. Upgrade the uderlay network protocol (IGP or BGP) to dual-stack to provide both IPv4 and IPv6 connectivity among the PEs/NVEs in the network. Both IPv4 and IPv6 underlay connectivity should be checked.

3. Upgrade EVPN VxLAN Encapsulation on the first VTEP (e.g., VTEP1) to dual-stack and configure IPv6 address for that VTEP in addition to its IPv4 address (e.g., dual-stack encapsulation).

4. The first VTEP (e.g., VTEP1) advertises EVPN routes with dual VTEP IP addresses, but the VXLAN encapsulation among all the VTEPs still uses "ipv4’", since the other VTEPs are still in single-stack "ipv4" mode.

5. Upgrade EVPN VxLAN Encapsulation on the second VTEP (e.g., VTEP2) to dual-stack and configure IPv6 address for that VTEP in addition to its IPv4 address (e.g., dual-stack encapsulation).

6. The second VTEP (e.g., VTEP2) advertises EVPN routes with dual VTEP IP addresses, and now the VXLAN encapsulation between VTEP1 and VTEP2 uses "dual-stack" and select "ipv6" as the prefered encapsulation. Section 3.2 describes the criteria used by a PE/NVE to prefer one encapsulation over another.

7. Upgrade EVPN VxLAN Encapsulation on the nth VTEP (e.g., VTEPn) to dual-stack and configure IPv6 address for that VTEP in addition to its IPv4 address (e.g., dual-stack encapsulation).

8. The nth VTEP (e.g., VTEPn) advertises EVPN routes with dual VTEP IP addresses, and now the VXLAN encapsulation between VTEP1 through VTEPn uses "dual-stack" and select "ipv6" as the prefered encapsulation. Section 3.2 describes the criteria used by a PE/NVE to prefer one encapsulation over another.

9. Once all the VTEPs have migrated to advertises EVPN routes with dual VTEP IP addresses and select "ipv6" as the preferred encapsulation and all BGP sessions in the network have also migrated to IPv6 (Section 3.4), the fabric has been migrated to use VXLANv6 in data plane with IPv6 BGP session in control, even though the all the VTEPs are still in "dual-stack" mode.

10. Furthermore, optionally the operator can set all the VTEPs to single-stack "ipv6" mode and advertise EVPN routes with single IPv6 VTEP IP address without impacting traffic, and then remove the IPv4 VTEP address on each of the VTEPs safely. Now the fabric has been migrated to use IPv6-only underlay.

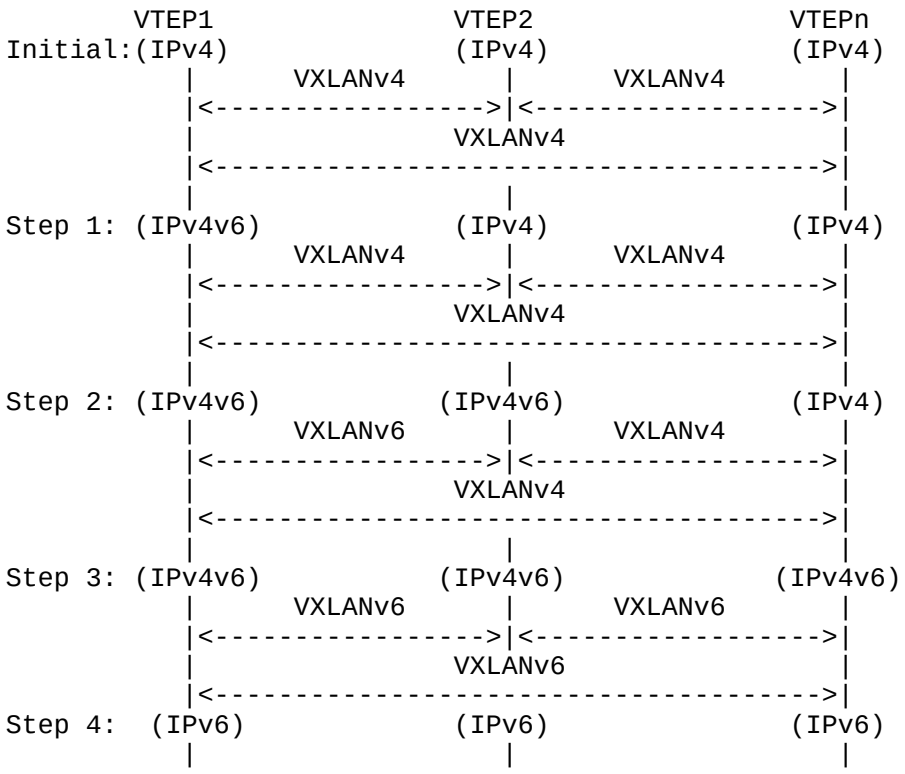

Figure 1: Fabric Migration from VXLANv4 to VXLANv6

Note: Migrating from Underlay IPv6 to IPv4 follows the same procedures as described above, as long as the originating router’s IP address doesn’t change during the migration process.

Note: Static Mulicast Underlay migration will be covered in future version.

3. BGP EVPN Routes and Procedures

3.1. Dual VTEP Address Placement

On a dual-stack underlay (XLANv4v6) VTEP, the EVPN routes advertised from the VTEP would carry both IPv4 and IPv6 VTEP addresses. The primary address would be carried in the Next Hop field of MP_REACH_NLRI attribute and secondary address is carried as a BGP Tunnel Encapsulation Attribute as defined in [RFC9012].

The Tunnel Encapsulation attribute is an optional transitive BGP path attribute. IANA has assigned the value 23 as the type code of the attribute in the "BGP Path Attributes" registry. As per section 6 in [RFC9012], the BGP Tunnel Encapsulation Attribute MAY be carried in any BGP UPDATE message whose AFI/SAFI is 25/70 (EVPN) and can be composed of a set of TLV encodings; however, for the purpose of dual-stack VxLAN Encapsulation, only a single Tunnel Encapsulation TLV is used and furthermore only a single sub-TLV for carrying VTEP IP address is used.

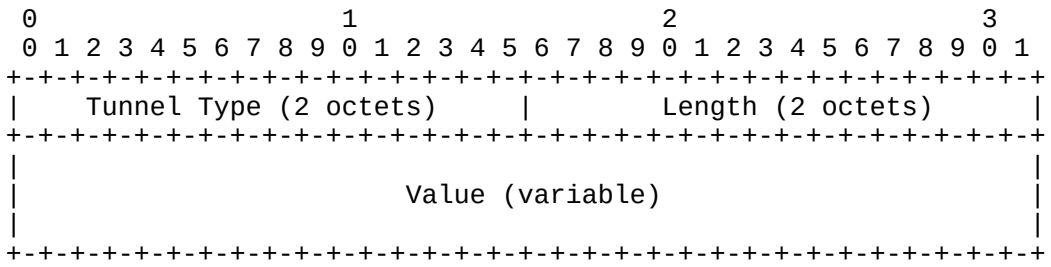

Figure 2: Tunnel Encapsulation TLV

As for VXLANv4v6 VTEP, within the Tunnel Encapsulation TLV, the Tunnel Type would be set to VXLAN (IANA assigned value 8) and the Tunnel Egress Endpoint sub-TLV field will carry the secondary IPv4 or IPv6 Next Hop Address.

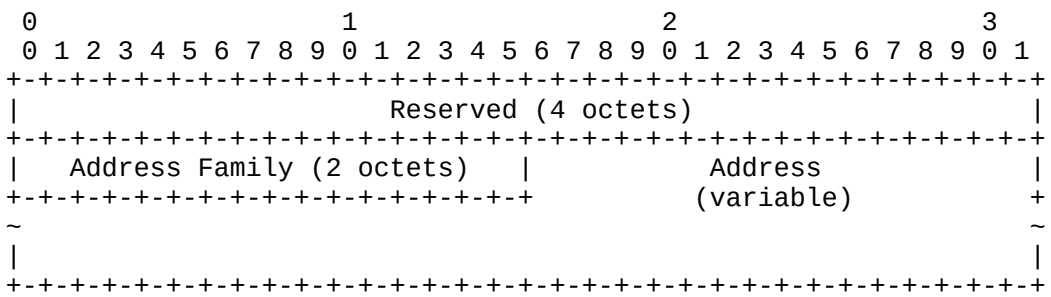

Figure 3: Tunnel Egress Endpoint sub-TLV Value Field

It should be noted that there is no need to carry the Encapsulation sub-TLV of 12 bytes in the Tunnel Encapsulation TLV because that info is already carried in the EVPN MAC/IP Advertisement Route. Furthermore, each EVPN route that carries the Tunnel Encapsulation Attribute MUST also carry BGP Encapsulation Extended Community per [RFC8365]. The presense of both the BGP Encapsulation Extended Community and the Tunnel Encapsulation Attribute indicates to the BGP EVPN receiver that the purpose of the Tunnel Encapsulation Attribute

is solely for carrying the second VTEP IP address and thus it only contains Tunnel Egress Endpoint sub-TLV (and doesn't contain any other sub-TLVs including VxLAN Encapsulation sub-TLV).

3.2. Dual VTEP Address Preference

For the interoperation between a dual-stack underlay VTEP and a single-stack underlay VTEP which doesn't support dual-stack underlay capability yet, a preference between choosing IPv4 or IPv6 as the primary address to be carried in the Next Hop field of BGP MP_REACH_NLRI attribute must be considered:

* For IPv4 to IPv6 fabric migration, a VXLANv4v6 VTEP must use its IPv4 VTEP address as the primary address and thus carry it in the MP_REACH_NLRI attribute. Its IPv6 VTEP address must be carried in Tunnel Encapsulation attribute.

* For IPv6 to IPv4 fabric migration, a VXLANv4v6 VTEP must use its IPv6 VTEP address as the primary address and thus carry it in the MP_REACH_NLRI attribute. Its IPv4 VTEP address must be carried in Tunnel Encapsulation attribute.

3.3. Originating Router's IP Address

As defined in the [RFC7432] and [RFC9251], there is an "Originating Router's IP Address" or "Originator Router Address" field in some EVPN routes, i.e., RT-3, RT-4, RT-6, RT-7, RT-8. This field can be 4 or 16 octets and it's part of the route key during the BGP route processing:

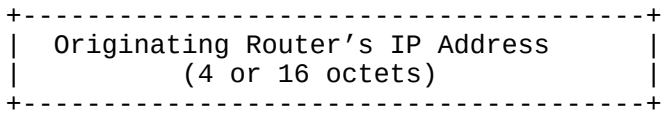

Figure 4: Originating Router's IP Address

Usually, for EVPN VXLANv4 routes, the Originating Router's IP Address can be assigned with an IPv4 loopback address, and for EVPN VXLANv6 routes, the Originating Router's IP Address can be assigned with an IPv6 loopback address.

Considering dual-stack underlay, if the originating router originally advertises a route with IPv4 underlay address, it might choose to use an IPv4 address as the "Originating Router's IP Address". Later on, when it tries to switch to the dual IPv4/IPv6 underlay address, it might advertise the route with an IPv6 address as the "Originating Router's IP Address". On the receiving side, since the two routes have different route keys, insteading of updating the previous route, both routes are kept which could lead to traffic duplication.

Considering that the "Originating Router's IP Address" is just part of the route key, which is used to uniquely identify an originating router, and it's not contributing to forwarding, it can be either IPv4 or IPv6 address independent of the BGP next hop address/VTEP address type for that NLRI, but it MUST remain the same for all EVPN routes advertised by that VTEP till the fabric migration is completed. These behaviours are clarified by section 8.1 and section 11.1 in [I-D.ietf-bess-rfc7432bis].

This implies that the originator of EVPN routes can use IPv4 address as "Originating Router's IP Address" while its encapsulation is VxLANv6, or it can use IPv6 address as "Originating Router's IP Address" while its encapsulation is VxLANv4. Further more, it implies that the EVPN receiver must accept either IPv4 or IPv6 as "Originating Router's IP Address" regardless of VxLAN encapsulation in dataplane.

3.4. BGP Peering

The BGP peering migration for BGP control plane signaling is independent from VTEP migration for dataplane connectivity and one can be done before the other independent of the order; however, before marking the fabric migration complete and and only have a single-stack enabled in the fabric, both dataplane and control plane migration needs to be completed.

For control plane migration (i.e., BGP session migration) from v4 to v6 (or visa versa), the following steps should be taken:

1. A given VTEP has a first BGP session (which is usually IPv4 for v4 to v6 migration) between itself and its BGP peer - i.e., either a route reflector in case of iBGP or another gateway/ASBR in case of eBGP.

2. A second BGP session using the other IP address type is configured between the VTEP and its peer (e.g., usually IPv6 address type for v4 to v6 migration). All EVPN routes get re-advertised using this new BGP session.

3. Once all the EVPN routes have been re-advertised via this second BGP session, then the first BGP session (which is IPv4 for v4 to v6 migration) can be decomissioned. Note that the same EVPN routes including tunnel attributes gets re-advertised via the two sessions.

Although as mentioned previously dataplane and control plane migrations from v4 to v6 (and vise versa) are independent of each other, it is recommended to perform dataplane migration first because as part of dataplane migration, the underlay connectivity among the network nodes (e.g., PE and P nodes) for the new IP address type is checked and this underlay connectivity is needed for control plane migration.

4. Reverting back to VxLANv4 before Migration Completion

Reverting the migration process back to VxLANv4 before its completion requires careful handling to ensure network stability and avoid traffic disruption. The following steps outline the procedure:

1. Determine which VTEPs have already been migrated to dual-stack mode and are advertising EVPN routes with both IPv4 and IPv6 VTEP addresses.

2. Start with the first dual-stack VTEP, revert its configuration to single-stack "ipv4" mode and ensure that the EVPN routes from the reverted VTEP are updated to reflect the single-stack "ipv4" configuration. The Next Hop field in the MP_REACH_NLRI attribute should only carry the IPv4 VTEP address. Do not unconfigure the IPv6 VTEP address at this moment to ensure seamless reverting.

3. Confirm that the reverted VTEP can communicate with other single-stack and dual-stack VTEPs in the fabric. Perform connectivity tests to ensure proper operation.

4. Repeat the above steps for each dual-stack VTEP in the fabric, one at a time, to ensure a controlled rollback process.

5. Once all dual-stack VTEPs have reverted to single-stack "ipv4" mode, all the VTEPs in fabric are using VXLANv4 in the data plane, and the IPv6 VTEP address can be safelly unconfigured.

By following these steps, the fabric can be reverted to VxLANv4 operation while maintaining network stability and minimizing impact on traffic.

5. Reverting back to VxLANv4 after Migration Completion

Reverting the migration process from VxLANv6 to VxLANv4 after its completion involves transitioning the fabric IPv6-only underlay back to IPv4-only underlay while ensuring minimal disruption to traffic. This is equivalent to fabric migration from IPv6 to IPv4 underlay. The following steps outline the procedure:

1. Confirm that all VTEPs in the fabric are operating in single-stack "ipv6" mode and are advertising EVPN routes with only IPv6 VTEP addresses.

2. Begin with the first VTEP, configure it to dual-stack mode by adding an IPv4 VTEP address alongside its existing IPv6 VTEP address. Ensure that the EVPN routes from this VTEP are updated to reflect the dual-stack configuration. The Next Hop field in the MP_REACH_NLRI attribute should carry the IPv6 VTEP address, while the IPv4 VTEP address is carried in the Tunnel Encapsulation attribute.

3. Confirm that the dual-stack VTEP can communicate with other single-stack and dual-stack VTEPs in the fabric. Perform connectivity tests to ensure proper operation.

4. Repeat the above steps for each VTEP in the fabric, transitioning them to dual-stack mode one at a time.

5. Once all VTEPs have transitioned to dual-stack mode, follow the same procedure as defined in Section 4 for the rest.

By following these steps, the fabric can be reverted to VxLANv4 operation after migration completion while maintaining network stability and minimizing impact on traffic.

6. Multi-Homing Operation

To support Single and Dual Stack EVPN VXLAN, the existing routes including multihoming related routes, need to be enhanced to support different Next Hops/Tunnel Endpoints (TEP) as defined in Section 3, but there is no changes to the migration procedures as described above.

6.1. RT-1 EAD Per EVI for Aliasing path

RT-1 is used together with RT-2 to resolve aliasing/backup path. Depending on the Preference, RT-1 and RT-2 could use different type of Next Hop. For example, RT-1 is from a Single IPv4 VTEP and RT-2 is from a Dual Stack VTEP, the two VTEPs form a multi-homing Group. A remote VTEP receives the RT-1 with IPv4 Next Hop and the RT-2 with Dual Next Hop, but locally prefer IPv6. In this case, the RT-2 would be resolved as one path with IPv4 Next Hop and another path with IPv6 Next Hop and the forwarding would still work.

6.2. RT-1 EAD Per ES for Fast Convergence

When a VTEP receives the RT-2's and builds the path list with primary and aliasing/backup paths, a shared next hop list will be built as well for fast convergence. As described above, the path list could be a mix of IPv4 and IPv6 Next Hops for the same Ethernet Segment. But since EAD Per ES route would share the same NH Preference as its corresponding EAD Per EVI and MAC/IP routes, the fast convergence would still work.

6.3. RT-1 EAD Per ES for Split Horizon Filtering

When a multihoming VTEP receives the EAD Per ES routes from its multihoming peers, it uses the VTEP Addresses from the routes to build the split horizon filtering list. Since EAD Per ES route would share the same NH Preference as its corresponding EAD Per EVI, MAC/IP and IMET routes, the same preferred VTEP address will be put into the filtering list, and the list could be a mix of IPv4 and IPv6 Next Hops for the same Ethernet Segment.

6.4. RT-4 ES Route for DF Election

ES Route is keyed with ESI and Originator's IP, and it's Next HOP Agnostic, so it would be safe to use either IPv4 or IPv6 as the Next Hop in MP_REACH_NLRI. Furthermore, an UPDATE to the NH of this route would not trigger DF Re-election.

6.5. RT-6, 7, 8 Routes for L2 Multicast

Like RT-4, the RT-6, 7 and 8 routes are Next Hop Agnostic, and use the Originator's IP Address as part of the route key. Since during migration the Originator's IP address won't change as described before, these L2 Multicast Routes won't be impacted by the VTEP IP/ Next Hop migration.

7. Fabric Interconnect between VXLANv4 and VXLANv6

[RFC9014] covers different interconnect networks for islands of EVPN VxLAN networks (i.e., NVO networks) among which EVPN MPLS and VxLAN networks. EVPN VxLAN as in interconnect network is covered in section xxx along with the corresponding procedures. When VxLAN encapsulation is used for both the NOV networks and the interconnect network, IP address types among the networks can be different without losing any generality in the procedures specified in [RFC9014]. Although both interconnect scenario as specified in [RFC9014] and fabric migration as specified in this document use dual-stack VTEPs, the application and procedures for the dual-stack VTEPs are different and are governed by their corresponding documents.

It should be noted that both interconnect and fabric migration scenarios can coexist such that several NVO networks are connected to an interconnect network, and two or more such networks (including interconnect network) have fabric migration from v4 to v6 (or vise versa). In such composite scenario, the procedures for each fabric migration are performed independently from the procedures for interconnect and the two sets of procedures do not interfer with each others.

8. Acknowledgements

The authors would like to thank Krishnaswamy Ananthamurthy, Krishna Deevi, Mohammed Mirza for their reviews of this document and feedbacks.

9. Security Considerations

All the security considerations in [RFC7432] apply directly to this document because this document leverages the control and data plane procedures described in those documents.

This document does not introduce any new security considerations beyond that of [RFC7432] because advertisements and processing of Ethernet Segment route for vES in this document follows that of physical ES in those RFCs.

10. IANA Considerations

This document requests no actions from IANA.

11. References

11.1. 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>.

[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>.

[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>.

[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>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <https://www.rfc-editor.org/info/rfc8365>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

[RFC9014] Rabadan, J., Ed., Sathappan, S., Henderickx, W., Sajassi, A., and J. Drake, "Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks", RFC 9014, DOI 10.17487/RFC9014, May 2021, <https://www.rfc-editor.org/info/rfc9014>.

[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>.

11.2. Informative References

[RFC7209] Sajassi, A., Aggarwal, R., Uttaro, J., Bitar, N., Henderickx, W., and A. Isaac, "Requirements for Ethernet VPN (EVPN)", RFC 7209, DOI 10.17487/RFC7209, May 2014, <https://www.rfc-editor.org/info/rfc7209>.

Authors' Addresses

Ali Sajassi
Cisco
Email: sajassi@cisco.com

Chuanfa Wang
Cisco
Email: chuanwan@cisco.com

Neeraj Malhotra
Cisco
Email: nmalhotr@cisco.com