Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 11 September 2026

J. Rabadan, Ed.
Nokia
A. Sajassi, Ed.
Cisco
E. Rosen
Individual
J. Drake
Independent
W. Lin
HPE
J. Uttaro
Independent
A. Simpson
Nokia

10 March 2026

Interconnecting EVPN and IPVPN Domains
draft-ietf-bess-evpn-ipvpn-interworking-18

Abstract

Ethernet Virtual Private Network (EVPN) provides a unified BGP control plane for both intra- and inter-subnet forwarding within tenant networks. When a tenant network spans multiple domains, including any combination of EVPN and IPVPN domains, it becomes necessary to define the interworking mechanisms among these BGP domains (EVPN and IPVPN) to ensure seamless end-to-end tenant connectivity. This document defines these interworking procedures.

In addition, this document defines a new BGP Path Attribute, referred to as D-PATH (Domain PATH), which provides loop prevention for gateway nodes by protecting against control plane loops. The introduction of D-PATH modifies the BGP best path selection process for Multiprotocol BGP inter-subnet forwarding routes of SAFI 128 (IPVPN) and SAFI 70 (EVPN).

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/.

Copyright Notice

Table of Contents

1. Introduction and Problem Statement

EVPN is used as a unified BGP control plane to support both intra-subnet and inter-subnet forwarding for tenant networks. In deployments where a tenant network spans multiple domains, some of which use EVPN, and others which rely on BGP VPN-IPv4/VPN-IPv6 address families for inter-subnet forwarding, it becomes necessary to define interworking procedures to enable seamless end-to-end tenant connectivity across these heterogeneous domains.

This document specifies procedures for interworking between EVPN and other BGP address families, including VPN-IPv4 and VPN-IPv6, for the purpose of inter-subnet forwarding. It also defines procedures for the interconnection of domains that may use EVPN, IPVPN, or a combination of both. Examples include the interconnection of two EVPN domains, two IPVPN domains, or an EVPN domain with an IPVPN domain.

To support loop prevention in scenarios where redundant gateway Provider Edges (PEs) interconnect distinct domains, this specification introduces a new BGP Path Attribute called the Domain Path (D-PATH). In topologies where multiple gateways connect domains, control plane loops may occur if routes are redistributed between domains without proper safeguards. For example, if gateway PE1 imports an IPVPN route for a given prefix and redistributes it as an EVPN IP Prefix route into the EVPN domain, and a second gateway PE2 receives this EVPN route and re-advertises it back into the IPVPN domain, a loop may form. The D-PATH attribute is designed to prevent such scenarios by providing domain-level loop detection and avoidance.

The D-PATH attribute alters the BGP best path selection logic for Multiprotocol BGP routes of SAFI 128 (VPN-IPv4/IPv6) and for EVPN IP Prefix routes. Accordingly, this document updates the BGP best path selection procedures specified in [RFC4271], but only for the IPVPN and EVPN families when the D-PATH attribute is used for inter-domain connectivity.

EVPN supports the advertisement of IPv4 or IPv6 prefixes through two route types:

* Route Type 2 - EVPN MAC/IP Advertisement route, as defined in [RFC9135], supporting host routes (i.e., /32 or /128).

* Route Type 5 - EVPN IP Prefix route, as defined in [RFC9136].

When interworking with other BGP address families for inter-subnet forwarding, the IP prefixes conveyed in these EVPN route types are re-originated into corresponding address families (e.g., IPVPN), and vice versa. Several aspects of this re-origination require clarified procedures, including route selection, loop prevention, and BGP Path Attribute handling across AFI/SAFI boundaries.

This document defines the concept of an Interworking PE (in Section 3), which is responsible for interconnecting different domains. An Interworking PE implements the following behavior: it imports routes from one domain (along with the domain-specific encapsulation parameters), installs them in an IP-VRF (IP Virtual Routing and Forwarding table [RFC9135]), and re-originates the routes with the encapsulation attributes suitable for the adjacent domain before advertisement. This reorigination process enables the solution to operate independently of the transport encapsulation mechanisms used within each domain and serves as a service interworking function.

The procedures defined herein ensure that tenant inter-subnet connectivity can be maintained across a mix of EVPN and non-EVPN domains, while preventing routing loops and maintaining protocol consistency across BGP address families.

As a summary, the following key procedures are specified by this document:

* A route selection algorithm that enables a PE to deterministically select the best path among candidates learned via EVPN and other ISF SAFIs.

* A new BGP Path Attribute, referred to as the Domain Path (D-PATH) attribute, which provides loop prevention capabilities and conveys domain traversal information for a given route.

* The rules governing BGP Path Attribute propagation across domains to maintain semantic consistency and enable cross-domain route processing.

* The operational procedures required on Interworking PEs that function as composite PEs, gateway PEs, or devices supporting both roles.

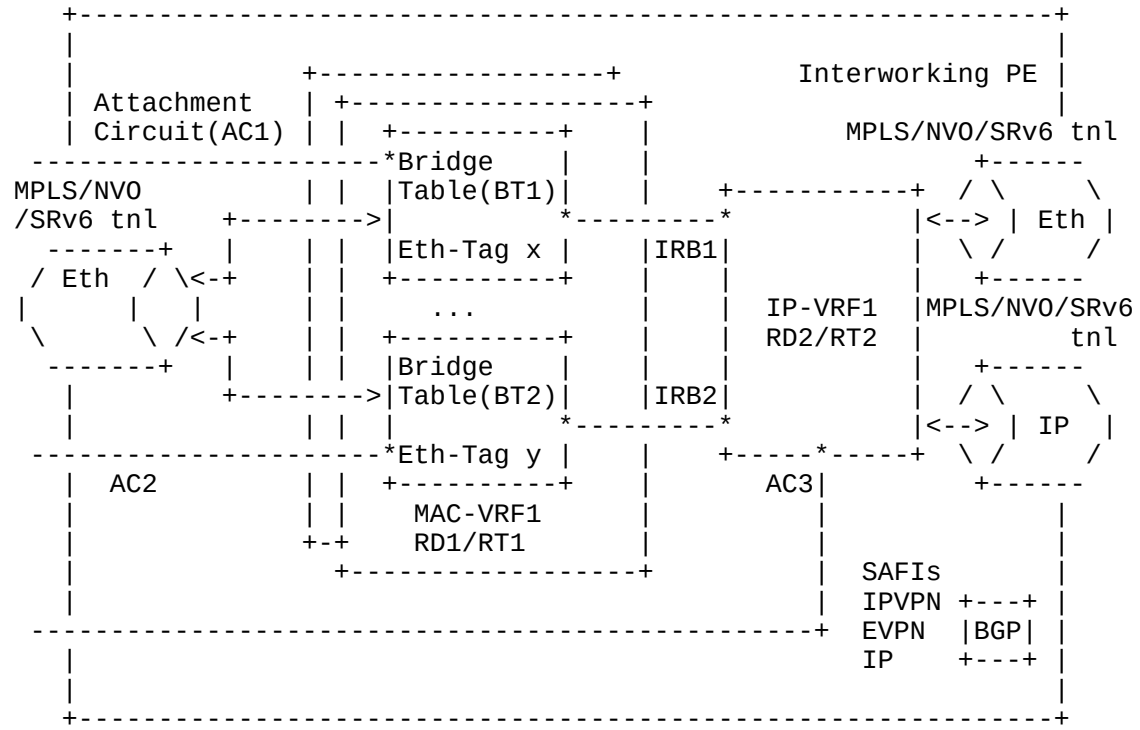
Collectively, these procedures equip operators with the necessary mechanisms to deploy scalable tenant networks spanning multiple administrative or routing domains, employing different ISF SAFIs for IP prefix dissemination while maintaining deterministic forwarding behavior and routing loop protection.

2. Conventions used in this document

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.

3. Terminology and Interworking PE Components

This section summarizes the terminology related to the "Interworking PE" concept that will be used throughout the rest of the document.



Note: tnl refer to "tunnel"

Figure 1: EVPN-IPVPN Interworking PE

* AC: Attachment Circuit or logical interface associated to a given BT or IP-VRF. To determine the AC on which a packet arrived, the PE will examine the combination of a physical port and VLAN tags (where the VLAN tags can be individual VLAN tags, Q-in-Q tags or ranges of both).

Example: In Figure 1, AC1 is associated to BT1, AC2 to BT2 and AC3 to IP-VRF1.

* BT: a Bridge Table, as defined in [RFC7432], represents the instantiation of a Broadcast Domain on a PE. When an EVI contains a single Broadcast Domain, the associated MAC-VRF on each PE includes a single BT. In cases where multiple Broadcast Domains exist within the same MAC-VRF, each BT is associated with a distinct Ethernet Tag. EVPN routes specific to a given BT include the corresponding Ethernet Tag to indicate the Broadcast Domain to which the route pertains.

  Example: In Figure 1, MAC-VRF1 has two BTs: BT1 and BT2. Ethernet Tag x is defined in BT1 and Ethernet Tag y in BT2.

* CE: Customer Edge device.

* Composite Domain: a domain in which multiple control plane ISF SAFIs, i.e., IPVPN and/or EVPN, are used and which is composed of regular PEs and composite PEs, see below.

* Composite PE: An Interworking PE that is connected to - at least - one composite domain and is capable of advertising a given prefix to multiple types of peers using appropriate route types. Specifically, a Composite PE advertises the prefix to an IPVPN peer using an IPVPN ISF route, to an EVPN peer using an EVPN ISF route, and to an RR (Route Reflector [RFC4456]) using both IPVPN and EVPN ISF routes (assuming the same RR is used for IPVPN and EVPN). A Composite PE implements the procedures defined in Section 7.

  Example: Figure 2 shows an example where PE1 is a composite PE since PE1 has EVPN and another ISF SAFI enabled to the same route-reflector, and PE1 advertises a given IP prefix IPn/x twice, one using EVPN and another one using ISF SAFI 128. PE2 and PE3 are not composite PEs.

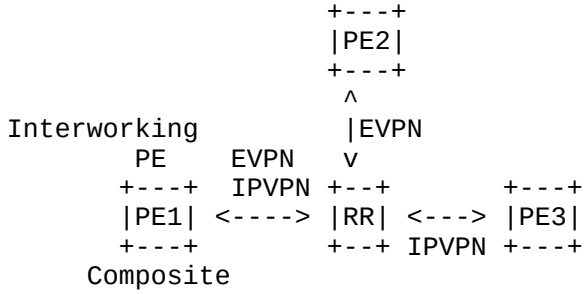


Figure 2: Interworking composite PE example

* Composite/Gateway PE: An Interworking PE that simultaneously performs the functions of both a Composite PE and a Gateway PE. This type of PE is connected to two or more domains: one (or more) regular domain and one (or more) composite domain. It operates as follows:

  - Re-originates an ISF route received from the regular domain into the composite domain. Within the composite domain, it performs the behavior of a Composite PE.

  - Re-originates an ISF route received from the composite domain into the regular domain. In the regular domain, the route is advertised using the ISF SAFI applicable to that domain.

  This functionality is particularly useful in scenarios where a tenant network spans multiple domains using different ISF SAFIs (e.g., IPVPN, and EVPN), and where any-to-any tenant connectivity is required. In such deployments, maintaining consistent end-to-end control plane behavior across domains is desirable when feasible.

  Example: Figure 3 illustrates an example where PE1 is a composite/gateway PE.

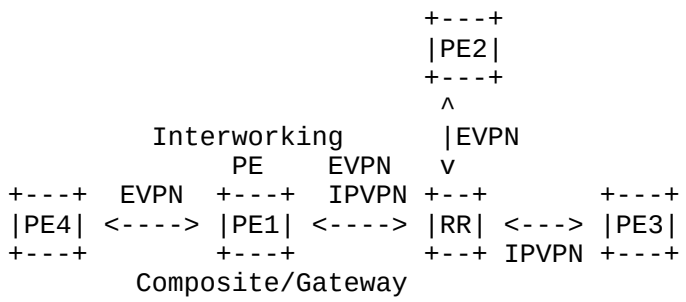


Figure 3: Interworking composite gateway PE example

* Domain: Two PEs belong to the same domain if they are attached to the same tenant and the packets exchanged between them do not require a data-path IP lookup (in the tenant space) at any transit router. A gateway PE interconnects multiple DOMAIN-IDs. Domain boundaries are not restricted to an Autonomous System or an IGP instance. The PEs in a domain may reside within the same or in different Autonomous Systems, and a single Autonomous System may also encompass multiple domains.

Example 1: Figure 4 depicts an example where Tenant Systems TS1 and TS2 belong to the same tenant, and they are located in different Data Centers that are connected by gateway PEs (see the gateway PE definition later). These gateway PEs use IPVPN in the WAN. When TS1 sends traffic to TS2, the intermediate routers between PE1 and PE2 require a tenant IP lookup in their IP-VRFs so that the packets can be forwarded. In this example there are three different domains. The gateway PEs connect the EVPN domains to the IPVPN domain.

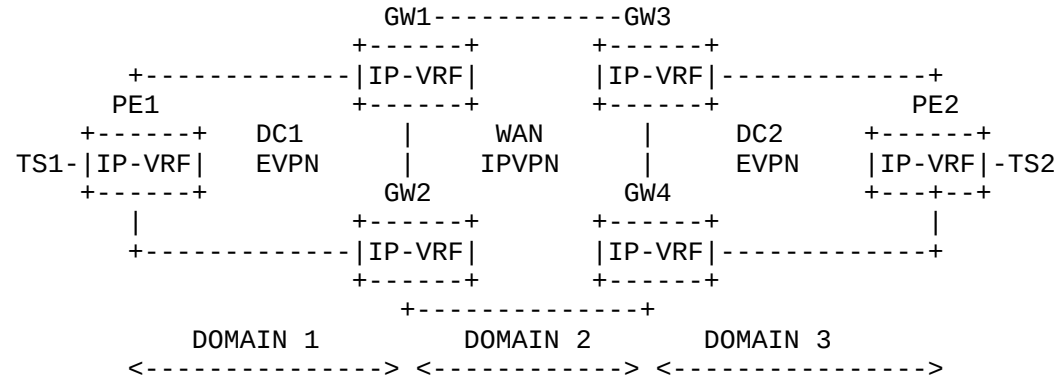


Figure 4: Multiple domain DCI example

Example 2: Figure 5 illustrates a similar example, but PE1 and PE2 are now connected by a BGP-LU (BGP Labeled Unicast) tunnel, and they have a BGP peer relationship for EVPN. Contrary to Example 1, there is no need for tenant IP lookups on the intermediate routers in order to forward packets between PE1 and PE2. Therefore, there is only one domain in the network and PE1/PE2 belong to it.

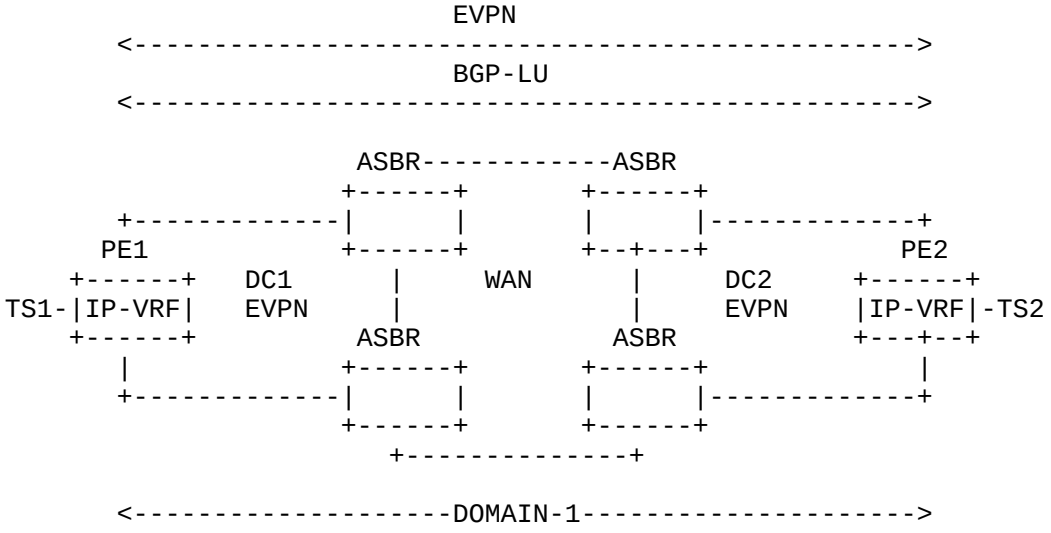


Figure 5: Single domain DCI example

* Ethernet Tag: used to represent a Broadcast Domain [RFC7432].

* EVI: an EVPN Instance spanning the Provider Edge devices participating in that EVPN [RFC7432].

* Gateway PE: An Interworking PE that connects two or more distinct domains, where each domain may be either a regular domain or a composite domain. A Gateway PE may establish either IBGP (Internal BGP [RFC4271]) or EBGP (External BGP [RFC4271]) sessions with peers in the connected domains. Depending on its configuration, the Gateway PE performs one of the following functions:

  - Re-originates ISF routes using the same ISF SAFI, between the connected domains.

  - Translates and re-originates an ISF route received with one ISF SAFI to a domain that uses a different ISF SAFI.

  A Gateway PE follows the procedures defined in Section 8. A gateway PE interconnects multiple domains. If the gateway PE is configured to use D-PATH, each domain is identified by a DOMAIN-ID and these DOMAIN-IDs are encoded in the D-PATH and are included in ISF SAFI route advertisements. The structure and behavior of the D-PATH attribute are described in Section 4.

Example: Figure 6 illustrates an example where PE1 is a gateway PE since the EVPN and IPVPN SAFIs are enabled on different BGP peers, and a given local IP prefix IPn/x is sent to both BGP peers for the same tenant. PE2 and PE1 are in one domain and PE3 and PE1 are in another domain.

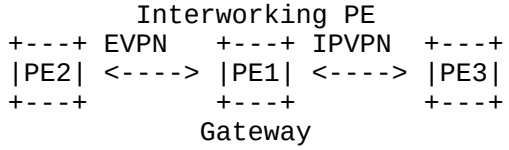


Figure 6: Interworking gateway PE example

* Interworking PE: A PE that is capable of advertising a given IP prefix using one or more of the following route types: an EVPN Inter-Subnet Forwarding (ISF) route, either an EVPN MAC/IP Advertisement route or an EVPN IP Prefix route, and an IPVPN ISF route. An Interworking PE maintains a single IP-VRF per tenant and zero, one, or more MAC-VRFs per tenant. Each MAC-VRF may include one or more Bridge Tables (BTs), and each BT may be associated with the tenant’s IP-VRF via an Integrated Routing and Bridging (IRB) interface. There are two types of Interworking PEs:

  - Composite PE

  - Gateway PE

  These two functions may be implemented independently on a per-tenant basis and may also coexist for the same tenant on a single PE.

  Example: Figure 1 shows an interworking PE, where ISF SAFIs are enabled. IP-VRF1 and MAC-VRF1 are instantiated on the PE, and together provide inter-subnet forwarding for the tenant.

* IP-VRF: an IP Virtual Routing and Forwarding table, as defined in [RFC4364][RFC9135]. Route Distinguisher and Route Target(s) are required properties of an IP-VRF. An IP-VRF is programmed with ISF routes.

* IRB: Integrated Routing and Bridging interface [RFC9135]. It refers to the logical interface that connects a BT to an IP-VRF and allows to forward packets with destination in a different subnet.

* ISF route: an Inter-Subnet Forwarding route for a given prefix, whose ISF SAFI may change as it transits different domains. IPVPN routes as in [RFC4364], [RFC4659], EVPN IP Prefix routes as in [RFC9136] or EVPN MAC/IP Advertisement routes when they are programmed within an IP-VRF [RFC9135], are considered ISF routes in this document.

* ISF SAFI: the Inter-Subnet Forwarding (ISF) Subsequent Address Family Identifier (SAFI) defines an MP-BGP (Multi Protocol Border Gateway Protocol [RFC4760]) Sub-Address Family used to advertise IP prefix reachability for inter-subnet forwarding within a tenant network. The SAFIs used for ISF include 1 (applicable only to IPv4 and IPv6 AFIs), 128 (applicable only to IPv4 and IPv6 AFIs), and 70 (EVPN, applicable only to AFI 25). The procedures defined in this document apply only to SAFI 128 and SAFI 70. Accordingly, for the purposes of this document, the term “ISF SAFI” refers exclusively to SAFI 128 or SAFI 70. The routes for these ISF SAFIs are referred to as IPVPN and EVPN routes. Note that the term “ISF SAFI” does not define a new SAFI; it is used solely as a collective reference to SAFI 128 and SAFI 70.

* MAC-VRF: a MAC Virtual Routing and Forwarding table, as defined in [RFC7432]. A MAC-VRF represents the instantiation of an EVPN Instance (EVI) on a PE. Each MAC-VRF is associated with a unique Route Distinguisher (RD) and one or more Route Targets (RTs), which are required attributes for its operation. These RD and RT values are typically distinct from those used by any associated IP-VRF, when such an IP-VRF is linked to the MAC-VRF through a Bridge Table via an Integrated Routing and Bridging (IRB) interface [RFC9135].

* MPLS/NVO/SRv6 tunnel: A tunnel that may be based on MPLS (Multi Protocol Label Switching) or a Network Virtualization Overlay (NVO) technology [RFC8365] or Segment Routing over IPv6 [RFC9252]. Such tunnels are utilized by both MAC-VRFs and IP-VRFs. Regardless of the underlying tunneling technology, the tunnel may carry either Ethernet or IP payloads. MAC-VRFs are restricted to using tunnels that carry Ethernet payloads - Ethernet NVO Tunnels [RFC9136], SRv6 tunnels with Ethernet payload - which are typically established via EVPN signaling. In contrast, IP-VRFs may utilize tunnels carrying Ethernet payloads, signaled via EVPN - or IP payloads, signaled via EVPN or IPVPN mechanisms. IPVPN-only PEs support IP-VRFs but do not support sending or receiving traffic over tunnels carrying Ethernet payloads.

  Example: Figure 1 illustrates the use of an MPLS, NVO-based or SRv6 tunnel to transport Ethernet frames associated with MAC-VRF1. The PE identifies the corresponding MAC-VRF and BT based on the

EVPN label - an MPLS label, a Virtual Network Identifier (VNI) or an SRv6 Segment ID -, depending on the encapsulation type. Additionally, Figure 1 shows two distinct MPLS/NVO/SRv6 tunnels used by IP-VRF1: one tunnel transports Ethernet frames, while the other carries IP packets. This demonstrates that IP-VRFs may concurrently utilize multiple tunnel types, depending on the payload and the signaling mechanism (EVPN or IPVPN).

* NVE: Network Virtualization Edge router [RFC8365].

* PE: Provider Edge device.

* Regular Domain: a domain in which a single control plane ISF SAFI, i.e., IPVPN or EVPN, is used. A Regular Domain is composed of regular PEs, see below. In Figure 4 and Figure 5, above, all domains are regular domains.

* Regular PE: A PE that is attached to a domain, either regular or composite, and which uses one of the control plane ISF SAFIs (IPVPN or EVPN) operating in the domain.

* RT-2: Route Type 2 or MAC/IP route, as per [RFC7432].

* RT-5: Route Type 5 or IP Prefix route, as per [RFC9136].

4. Domain Path Attribute (D-PATH)

The BGP D-PATH attribute is an optional and transitive BGP path attribute.

Similar to AS_PATH, D-PATH is composed of a sequence of Domain segments. Each Domain segment is composed of <domain segment length, domain segment value>, where the domain segment value is a sequence of one or more Domains, as illustrated in Figure 7. Each domain is represented by <DOMAIN-ID:ISF_SAFI_TYPE>.

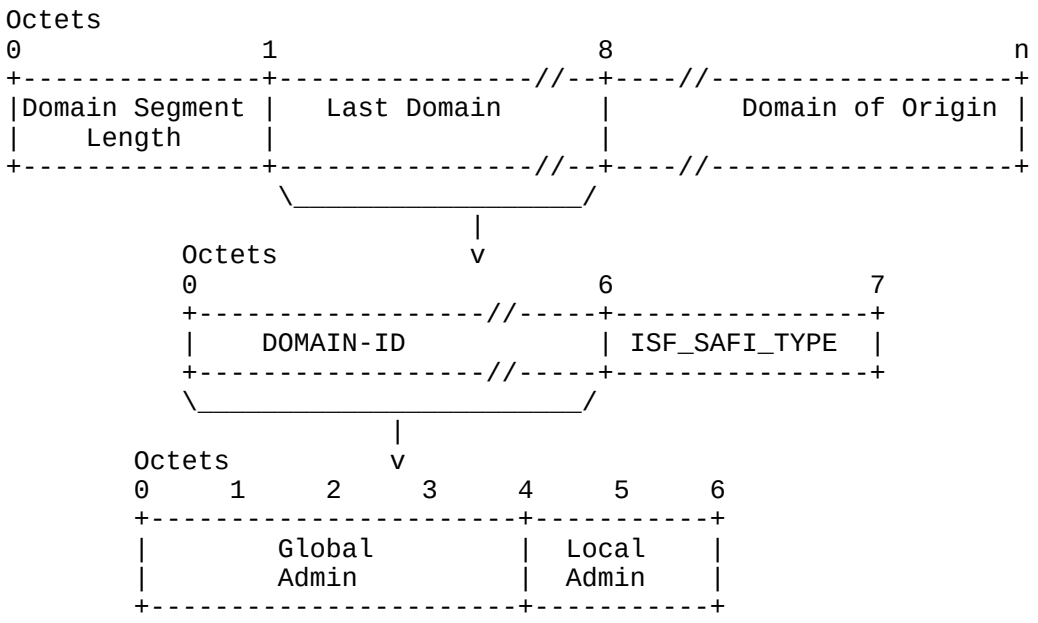


Figure 7: D-PATH Domain Segment

* Domain Segment Length (length: 1-octet): containing the number of domains in the segment.

* “Last Domain” refers to the most recently added Domain, while “Domain of Origin” refers to the first Domain added by the gateway PE that initialized the D-PATH for the ISF route. Multiple Domains may exist between those Domains.

* DOMAIN-ID is a 6-octet field that represents a domain. It is composed of a 4-octet Global Administrator sub-field and a 2-octet Local Administrator sub-field. The Global Administrator sub-field MAY be filled with an Autonomous System Number (ASN, Public or Private), an IPv4 address, or any value. The combined Global Administrator and Local Administrator can use any value that guarantees the uniqueness of the DOMAIN-ID (when the tenant network is connected to multiple Operators) and helps troubleshooting and debugging of D-PATH in ISF routes. A Gateway PE that interconnects two domains is associated with two distinct DOMAIN-IDs, one per domain. All Gateway PEs attached to the same domain MUST use the same DOMAIN-ID value to represent that domain. Expressing the Global Administrator and Local Administrator values as opaque unsigned integers in user interface and reporting (e.g., CLI/YANG) is RECOMMENDED.

* ISF_SAFI_TYPE is a 1-octet field that indicates the Inter-Subnet Forwarding SAFI type in which a route was received by the gateway PE, before the route is re-exported by the gateway PE into a different domain. The ISF_SAFI_TYPE field is informational and does not have any impact on the loop detection or BGP Path selection procedures. Encoding the ISF_SAFI_TYPE provides operational benefits, as it allows operators to verify that the intended interworking is in place and that the route has traversed the expected domains using the intended ISF SAFIs in each domain. The non-zero ISF_SAFI_TYPE values come from the IANA SAFI registry [IANA-SAFI]. These are the values allowed by this document:

Value	ISF_SAFI_TYPE
0	Gateway PE local ISF route
70	EVPN
128	IPVPN

Table 1

The BGP D-PATH attribute is supported on ISF routes of type IPVPN and EVPN and MUST NOT be advertised along with routes different from IPVPN and EVPN routes. By default, the BGP D-PATH attribute is not advertised and MUST be explicitly enabled by configuration on the Gateway PEs. The rest of this section specifies the D-PATH related procedures:

a. D-PATH identifies the sequence of domains, each identified by a <DOMAIN-ID:ISF_SAFI_TYPE> through which a given ISF route of type IPVPN or EVPN has passed.

    * This attribute list MAY contain one or more segments. Each segment’s Domain Segment Length MUST be equal or greater than one.

    * The first entry in the list (leftmost) is the <DOMAIN-ID:ISF_SAFI_TYPE> from which a gateway PE is re-originating an ISF IPVPN or EVPN route. The last entry in the list (rightmost) is the <DOMAIN-ID:ISF_SAFI_TYPE> from which a gateway PE received an ISF IPVPN or EVPN route without a D-PATH attribute (the Domain of Origin). Intermediate entries in the list are domains that the ISF IPVPN or EVPN route has transited.

* As an example, an ISF IPVPN or EVPN route received with a D-PATH attribute containing a domain segment of {length=2, <6500:2:IPVPN>,<6500:1:EVPN>} indicates that the route was originated in EVPN domain 6500:1, and re-originated into IPVPN domain 6500:2.

* In order to minimize the number of segments in the D-PATH attribute, the local gateway PE MUST prepend its own domain as the last element of the domain segment. If the act of prepending a new domain causes an overflow in the domain segment (i.e., more than 255 domains), the local gateway PE MUST prepend a new segment and prepend its own domain to this new segment.

b. D-PATH is added/modified by a gateway PE when re-originating an update to a different domain (which runs the same or different ISF SAFI), assuming the use of D-PATH is configured:

* The IP-VRF of a Gateway PE that interconnects two domains is associated with two distinct DOMAIN-IDs, one per domain. These DOMAIN-IDs MUST be different. Each domain MUST be identified by a unique DOMAIN-ID. All Gateway PEs attached to the same domain MUST use the same DOMAIN-ID value to represent that domain.

* Whenever a prefix arrives at a gateway PE in a particular ISF SAFI route, if the gateway PE needs to export that prefix to a BGP peer, the gateway PE MUST prepend a <DOMAIN-ID:ISF_SAFI_TYPE> to the list of domains in the D-PATH of the received route, as long as the gateway PE works in Uniform-Propagation-Mode, as explained in Section 5.2, and the use of D-PATH is configured as described at the beginning of this section.

* For instance, consider an IP-VRF configured with DOMAIN-IDs 6500:1 for EVPN and 6500:2 for IPVPN. If an EVPN route for prefix P is received and P is installed in the IP-VRF, then the corresponding IPVPN route for P, when exported to an IPVPN peer, will include the domain identifier <6500:1:EVPN> prepended to the existing D-PATH attribute, assuming the use of D-PATH is configured, as described at the beginning of this section. Similarly, prefixes received in the IP-VRF from an IPVPN peer will be exported to EVPN peers with the domain identifier <6500:2:IPVPN> appended to the D-PATH attribute, again assuming the use of D-PATH is configured.

* In the above example, if the EVPN route is received without D-PATH, the gateway PE will add the D-PATH attribute with one segment {length=1, <6500:1:EVPN>} when re-advertising to domain 6500:2.

* Within the Domain of Origin, the update does not contain a D-PATH attribute because the update has not passed through a gateway PE yet.

c. For a local ISF route, i.e., a configured static route or a route learned from a local attachment circuit, a gateway PE following this specification has three choices:

1. The gateway PE advertises that ISF route without a D-PATH attribute into one or more of its configured domains, in which case the D-PATH attribute will be added by the other gateway PEs in each of those domains.

2. The gateway PE advertises that ISF route with a D-PATH attribute into one or more of its configured domains (assuming the use of D-PATH is configured), in which case the D-PATH attribute in each copy of the ISF route is initialized with an ISF_SAFI_TYPE of 0 and the DOMAIN-ID of the domain with which the ISF route is associated.

3. The gateway PE advertises the ISF route with a D-PATH attribute (assuming the use of D-PATH is configured) containing a locally configured domain identifier associated with its local ISF routes into one or more of its configured domains. In this case, the D-PATH attribute in each copy of the ISF route is initialized with an ISF_SAFI_TYPE value of 0 and the DOMAIN-ID representing the local ISF domain. The DOMAIN-ID MUST be globally unique and MAY be shared across multiple gateway PEs.

   Although all three options provide mechanisms for detecting control plane loops, this third option is RECOMMENDED, as it conveys additional information about the origin of the route. Specifically, it allows the receiving PE to identify the route as having originated from a local gateway, based on the combination of the DOMAIN-ID and the ISF_SAFI_TYPE value.

d. An ISF route of type IPVPN or EVPN received by a Gateway PE that includes a D-PATH attribute containing one or more DOMAIN-ID values locally associated with the corresponding IP-VRF MUST be considered a looped ISF route for the purposes of re-advertisement into adjacent domains. In such cases:

* The ISF route MUST be flagged as "looped".

* The route MUST NOT be re-exported to any other domain.

* The route is installed in the IP-VRF only if it is selected as the best path according to the procedures defined in Section 6.

For the purpose of loop detection, the ISF_SAFI_TYPE value associated with a DOMAIN-ID in the D-PATH attribute is irrelevant. That is, a route is considered looped if it contains at least one DOMAIN-ID that matches any local DOMAIN-ID configured on the Gateway PE, regardless of the ISF_SAFI_TYPE value.

Example: In the scenario illustrated in Figure 4, gateway GW1 receives two ISF routes for the same prefix associated with TS1:

* An EVPN IP Prefix route with a next-hop of PE1, and no D-PATH attribute.

* An IPVPN route with a next-hop of GW2, and a D-PATH attribute containing a single segment: {length=1, <6500:1:EVPN>}, where 6500:1 is assumed to be the DOMAIN-ID for domain 1, which is local to GW1.

Upon receiving the IPVPN route, GW1 identifies 6500:1 as a locally configured DOMAIN-ID, and therefore flags the route as "looped". As a result, GW1 does not install this route in the tenant IP-VRF, because the route selection process prefers the EVPN IP Prefix route (due to its shorter D-PATH attribute, as specified in Section 6). Loop detection is applied even if the ISF_SAFI_TYPE value in the D-PATH attribute is unknown to GW1 or does not match any SAFI defined in this specification.

e. A DOMAIN-ID configured on a gateway PE MAY be assigned at either the domain interconnection level or scoped individually per tenant IP-VRF.

* When the DOMAIN-ID is allocated at the peering domain level, it SHALL apply to all tenant IP-VRFs associated with that domain.

* When the DOMAIN-ID is allocated for a specific tenant IP-VRF, the processing of received D-PATH attributes and their subsequent propagation SHALL be performed in the context of that IP-VRF’s DOMAIN-ID.

A per tenant IP-VRF DOMAIN-ID assignment is particularly useful in scenarios involving route leaking. For example, consider two gateway PEs, PE1 and PE2, both associated with different tenant IP-VRFs, denoted as IP-VRF-1 and IP-VRF-2. If PE1 advertises ISF SAFI routes for IP-VRF-1 with a DOMAIN-ID of 6500:1, and these routes are received on PE2 and subsequently leaked from IP-VRF-1 into IP-VRF-2, the re-advertisement of the routes from PE2 back to PE1 in the context of IP-VRF-2 will not be considered looped by PE1. This is because PE1 processes the route in the context of IP-VRF-2, for which DOMAIN-ID 6500:1 is not locally configured.

f. The number of domains encoded in the D-PATH attribute reflects the number of Gateway PEs that the corresponding ISF route update has traversed. If a transit Gateway PE performs route leaking between two local tenant IP-VRFs, it MAY prepend a domain to the D-PATH attribute with an ISF_SAFI_TYPE value of 0 when exporting the leaked route into an ISF SAFI. In such cases, the total number of domain entries in the D-PATH attribute reflects not only the number of Gateway PEs through which the ISF route has been re-originated, but also the number of tenant IP-VRF instances across those Gateway PEs.

g. The following error-handling procedures apply to the D-PATH Path Attribute:

   1. A received D-PATH attribute MUST be considered malformed if it contains a malformed Domain Segment or if the total length of the D-PATH attribute is less than eight octets.

   2. A Domain Segment MUST be considered malformed under any of the following conditions:

      * The length of the Domain Segment is zero.

      * The length of the Domain Segment exceeds the remaining length of the enclosing D-PATH attribute.

      * Fewer than eight octets remain after the last successfully parsed Domain Segment.

      * Each Domain Segment consists of a one-octet length field indicating the number of Domains in the segment, with each Domain encoded in seven octets. If the total length of the Domain Segment (i.e., 1 + 7 × number of Domains) exceeds the remaining length of the D-PATH attribute, the Domain Segment is considered malformed.

3. A BGP speaker receiving an UPDATE message containing a malformed D-PATH attribute SHALL apply the "treat-as-withdraw" procedure, as specified in [RFC7606].

4. Domains within the D-PATH attribute that contain unrecognized ISF_SAFI_TYPE values MAY be accepted and MUST NOT be considered an error.

5. The D-PATH Path Attribute MUST NOT appear more than once in the Path Attributes of a given BGP UPDATE message. If multiple instances of the D-PATH attribute are present, all instances other than the first MUST be discarded, and the UPDATE message MUST continue to be processed. This behavior follows [RFC7606], including the associated logging considerations.

6. The D-PATH Path Attribute MAY be included only in UPDATE messages that carry IPVPN or EVPN routes. It MUST NOT be included with any other AFI/SAFI combinations. If a D-PATH attribute is received in an UPDATE message associated with an unsupported AFI/SAFI, the "treat-as-withdraw" procedure MUST be applied, in accordance with [RFC7606].

5. BGP Path Attribute Propagation across Domains

A Gateway PE, depending on its local configuration, is required to re-originate an ISF route between two domains that utilize either the same or different ISF SAFIs. This requires defining how a Gateway PE handles the BGP Path Attributes associated with the ISF route during such re-origination.

This section specifies the BGP Path Attribute propagation behaviors that a Gateway PE MAY apply when it receives an ISF route with ISF SAFI x, installs the route into the relevant IP-VRF, and subsequently re-advertises the route as an ISF route using ISF SAFI y. The values of ISF SAFI x and SAFI y MAY be the same or different.

5.1. No-Propagation Mode

The No-Propagation Mode is the default operational mode for Gateway PEs when re-exporting ISF routes from one domain into another. In this mode, the Gateway PE re-initializes the BGP Path Attributes during the re-origination of an ISF route, treating it in the same manner as a directly connected or locally originated IP prefix.

This mode is suitable for deployment scenarios where the source domain - for example, an EVPN domain - is "abstracted" and treated as a virtual CE, and where remote IPVPN or IP-based PEs do not rely on the BGP Path Attributes of the source EVPN domain for best-path selection or the application of routing policy.

It is important to note that, in No-Propagation Mode, the D-PATH attribute is not propagated. As a result, redundant Gateway PEs may be susceptible to routing loops. While such loops may be mitigated using routing policies or additional attributes, such as the Route Origin extended community [RFC4360], this approach does not guarantee detection or prevention of all potential loop scenarios.

5.2. Uniform Propagation Mode

In Uniform Propagation Mode, the Gateway PE retains and copies a consistent set of commonly used BGP Path Attributes when re-originating an ISF route between domains. This mode is typically employed in deployments where IP prefixes are seamlessly distributed using both EVPN and/or IPVPN SAFIs. This specification permits the propagation of a limited set of commonly used attributes, while discouraging indiscriminate copying and re-advertisement, primarily for security reasons.

The following normative behavior MUST be followed by a Gateway PE operating in Uniform Propagation Mode:

1. Upon receiving an ISF route, and provided that no validation errors are detected and the route is permitted by local policy, the gateway PE imports the route into the associated IP-VRF and retains the original BGP Path Attributes. When re-advertising the route into a different domain, the gateway PE SHOULD, by default, propagate only the following set of attributes. All other Path Attributes SHOULD NOT be propagated unless explicitly permitted by local import/export policies:

    * AS_PATH

    * D-PATH (only when advertising IPVPN or EVPN routes)

    * IBGP-only attributes (when advertising to IBGP peers): LOCAL_PREF, ORIGINATOR_ID, CLUSTER_ID

    * MULTI_EXIT_DISC (MED)

    * AIGP [RFC7311]

* COMMUNITY, EXTENDED_COMMUNITY, and LARGE_COMMUNITY, except where explicitly excluded in Item 4 below.

2. When re-advertising an ISF route to an IBGP peer, the gateway PE SHOULD preserve the AS_PATH of the original ISF route without modification. When re-advertising to an EBGP peer, the Gateway PE SHOULD prepend the IP-VRF's ASN to the preserved AS_PATH.

3. When re-originating an ISF route to IBGP peers, the gateway PE SHOULD retain IBGP-only attributes (e.g., LOCAL_PREF, ORIGINATOR_ID, CLUSTER_ID) from the original ISF route. As the route is re-originated, the gateway PE is not required to perform the route reflector function described in [RFC4456].

4. As stated in Item 1, the gateway PE SHOULD preserve the COMMUNITY, EXTENDED_COMMUNITY, and LARGE_COMMUNITY attributes from the original ISF route. However, the following exceptions apply:

   a. BGP Encapsulation Extended Communities, as defined in [RFC9012], SHOULD NOT be propagated.

   b. Route Target Extended Communities SHOULD NOT be propagated and SHOULD be re-initialized when re-advertising the ISF route into a different domain. The re-initialized Route Target value MAY match the value used in the original route.

   c. All EVPN-specific Extended Communities SHOULD NOT be propagated.

   d. Gateway PEs SHOULD support import/export policies capable of matching COMMUNITY, EXTENDED_COMMUNITY, and LARGE_COMMUNITY values to permit or deny their propagation between domains when the default propagation behavior needs to be overridden.

   The Gateway PE SHOULD NOT copy the above Extended Community types in "a", "b" and "c" from the original ISF route into the re-advertised ISF route. Certain Extended Communities may influence how the receiving PE processes the route. Propagating such attributes into another domain could therefore lead to unintended behavior. For example, if the BGP Encapsulation Extended Community is propagated into a destination domain that uses a different encapsulation, a receiving PE in that domain might interpret the label field of the EVPN ISF route according to an encapsulation context that does not apply locally [RFC8365]. This could result in the route being discarded or programmed with incorrect encapsulation parameters.

5. For a given ISF route, only the BGP Path Attributes associated with the best path MAY be propagated when re-advertising the route into a different domain. If multiple paths are received for the same prefix within the same ISF SAFI, the standard BGP best path selection procedure MUST be applied to determine the active path and its associated attributes. Even when Equal-Cost Multi-Path (ECMP) is enabled for the IP-VRF, only the Path Attributes of the selected best path SHOULD be propagated.

5.3. Aggregation of Routes and Path Attribute Propagation

Instead of re-originating a high number of (host) ISF routes between domains, a gateway PE that receives multiple ISF routes from a domain MAY choose to re-originate a single ISF aggregate route into a different domain. In this document, aggregation is used to combine the characteristics of multiple ISF routes in such way that a single aggregate ISF route can be re-originated to the destination domain. Aggregation of multiple ISF routes of one ISF SAFI into an aggregate ISF route is only done by a gateway PE.

Aggregation on gateway PEs may use either the No-Propagation-Mode or the Uniform-Propagation-Mode explained in Section 5.1 and Section 5.2, respectively.

When using Uniform-Propagation-Mode, Path Attributes of the same type code MAY be aggregated according to the following rules:

* AS_PATH is aggregated based on the rules in [RFC4271]. The gateway PEs are not expected to receive AS_PATH attributes with path segments of type AS_SET [RFC9774]. Routes received with AS_PATH attributes including AS_SET path segments MUST NOT be aggregated.

* An ISF aggregate route SHOULD NOT be advertised unless all the contributing ISF routes have the same D-PATH DOMAIN-ID members, regardless of their order. If there is at least one contributing ISF route that has a different D-PATH DOMAIN-ID, the gateway PE SHOULD advertise each contributing ISF route with its own D-PATH (prepended with the gateway’s domain). An implementation MAY, by local policy, override this behavior and advertise an ISF aggregate route without the D-PATH attribute when the contributing routes do not share identical D-PATH DOMAIN-ID members. In such cases, redundant gateway PEs SHOULD apply a consistent policy to prevent the advertisement of aggregate routes with inconsistent D-PATH usage into the destination domain.

* The Community, Extended Community and Large Community attributes of an aggregated ISF route SHOULD include the union of the corresponding attributes from all constituent ISF routes that were aggregated, with the exception of those Extended Community types explicitly excluded from propagation as specified in Section 5.2, or those for which the applicable specifications define different handling.

* For other attributes, rules in [RFC4271] or the attribute applicable specifications are followed.

If the conditions for route aggregation, as specified above, are satisfied, operators SHOULD consider enabling aggregation in environments with large-scale tenant networks where a significant number of host routes are present. This practice is particularly applicable to deployments such as large-scale data centers.

6. Route Selection Process for ISF Routes

A PE router may receive the same IP prefix via ISF routes with different ISF SAFIs, and from either the same or different BGP peers. Additionally, the same IP prefix (e.g., a host route) may be received in both an EVPN MAC/IP Advertisement route and an EVPN IP Prefix route. To ensure consistent and deterministic forwarding behavior, a route selection procedure across all ISF SAFIs is required.

The objectives of this route selection process are as follows:

* To ensure that all composite and gateway PEs have a consistent and deterministic view of the preferred path to reach a given IP prefix.

* To enable meaningful comparison of routes advertised in EVPN and non-EVPN ISF SAFIs based on commonly used path attributes.

* To support Equal-Cost Multi-Path (ECMP) forwarding across EVPN and non-EVPN ISF SAFI routes, where applicable.

For a given prefix received via one or more non-EVPN ISF routes, the standard BGP best path selection procedure, as defined in [RFC4271], is applied to determine the "non-EVPN best paths." Similarly, for a given prefix received via one or more EVPN ISF routes, the same procedure is applied to determine the "EVPN best paths."

When both EVPN and non-EVPN ISF routes are present for the same prefix within a single IP-VRF, the PE MUST perform a tie-breaking selection procedure on the union of these best-path sets. The process treats all candidate ISF routes as equally preferable initially, then iteratively removes routes until a single best path (or a valid ECMP set) remains.

6.1. Tie-Breaking and Selection Rules

The selection procedure MUST follow the standard route selection rules defined in [RFC4271], with the following additional rules and exceptions applied in the specified order:

1. Immediately after applying the Local Preference comparison step from [RFC4271], the PE MUST remove from consideration any routes that do not have the shortest D-PATH attribute. Routes with no D-PATH attribute are considered to have a D-PATH length of zero. This rule MUST NOT be applied to ISF routes that are not imported into an IP-VRF.

2. After applying Rule 1, the standard [RFC4271] selection steps MUST continue in order.

3. If, after the previous steps, one or more candidate routes remain and at least one of them is an EVPN MAC/IP Advertisement route (EVPN Route Type 2), then all EVPN IP Prefix routes (EVPN Route Type 5) MUST be removed from consideration.

4. If ECMP is enabled by policy and the remaining candidate routes after Steps 1 through 3 include both EVPN and non-EVPN paths, then both paths MUST be retained. If ECMP is not enabled, and such a case arises, the EVPN path MUST be selected and the non-EVPN path MUST be removed from consideration.

This procedure extends the standard BGP best path selection behavior as specified in [RFC4271] for IPVPN and EVPN IP Prefix routes by incorporating D-PATH based tie-breaking to prefer routes that traverse the fewest Gateway PEs or domains. These rules MUST NOT be applied to routes received under AFI/SAFI combinations other than IPVPN or EVPN; such routes - different from IPVPN or EVPN - get treat-as-withdraw procedures if they are received with a D-PATH attribute, as described in Section 4.

6.2. Examples

Example 1:

PE1 receives three candidate routes for prefix IP1/32, all eligible for import into IP-VRF-1:

{SAFI=EVPN, RT-2, Local-Pref=100, AS-Path=(65536,65537)}
{SAFI=EVPN, RT-5, Local-Pref=100, AS-Path=(65536,65537)}
{SAFI=128, Local-Pref=100, AS-Path=(65536,65537)}

Selected route:

{SAFI=EVPN, RT-2, Local-Pref=100, AS_PATH=(65536,65537)}

This outcome is due to Step 3, which gives preference to Route Type 2 when both Type 2 and Type 5 EVPN routes exist.

Example 2:

PE1 receives two candidate routes for prefix IP2/24, both eligible for import into IP-VRF-1:

{SAFI=EVPN, RT-5, D-PATH=(6500:3:IPVPN), AS-Path=(65536,65537), MED=10}
{SAFI=128, D-PATH=(6500:1:EVPN,6500:2:IPVPN), AS-Path=(65537), MED=200}

Selected route: {SAFI=EVPN, RT-5, D-PATH=(6500:3:IPVPN), AS_PATH=(65536,65537), MED=10}

This result is due to Step 1, which prefers the route with the shortest D-PATH.

7. Composite PE Procedures

As described in Section 3, composite PEs are typically used in tenant networks where EVPN and IPVPN are both used to provide inter-subnet forwarding within the same composite domain.

Figure 8 depicts an example of a composite domain, where PE1/PE2/PE4 are composite PEs (they support EVPN and IPVPN ISF SAFIs on their peering to the Route Reflector), and PE3 is a regular IPVPN PE.

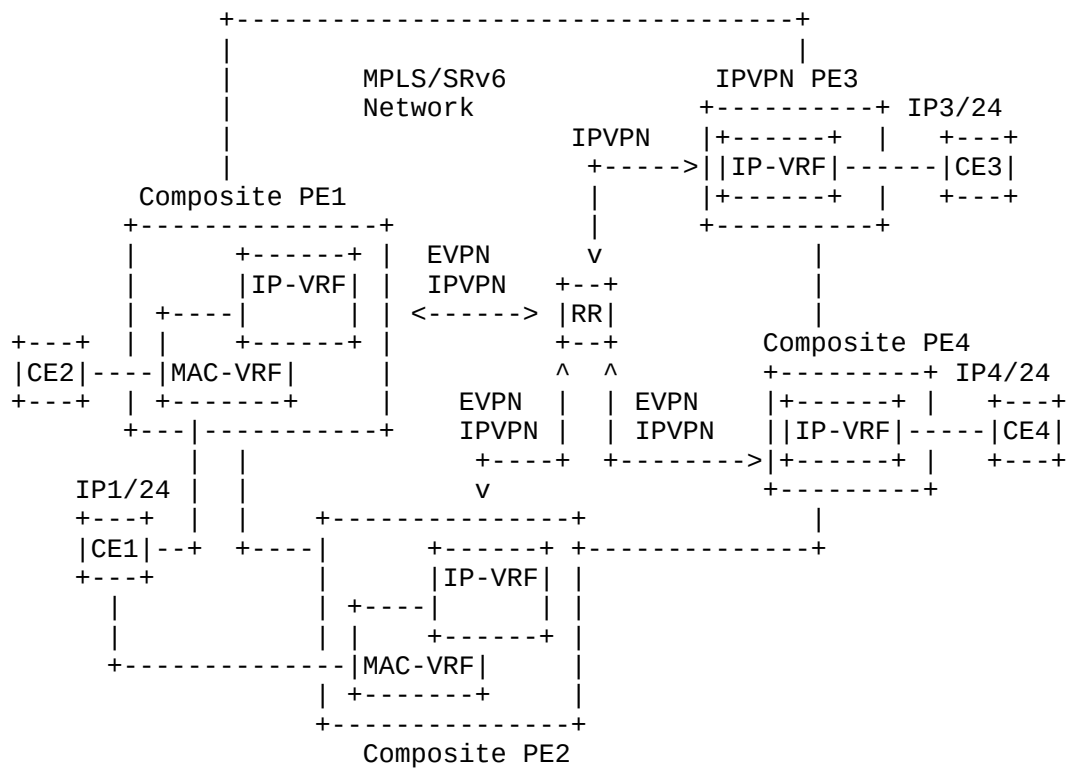


Figure 8: Composite PE example

In a composite domain comprising both composite and regular PEs, the following behaviors apply:

1. Prefix Advertisement Consistency

   Composite PEs MUST advertise the same IP prefixes using each ISF SAFI to the Route Reflector (RR), assuming the same RR is used for both ISF SAFIs. For example, as shown in Figure 8, the prefix IP1/24 is advertised by PE1 and PE2 to the Route Reflector in two separate NLRI entries: one for AFI/SAFI 1/128 (IPVPN) and another for EVPN. If both routes are advertised with the same set of BGP Path Attributes, the receiving composite PE will select the EVPN route over the IPVPN route, following the route selection procedures defined in Section 6. Prioritizing the advertisement of the EVPN route before the IPVPN route is an OPTIONAL optimization. This ensures that the EVPN route is more likely to be selected first, avoiding unnecessary replacement if the IPVPN route arrives later.

2. Route Reflector SAFI-Specific Forwarding Behavior

   The Route Reflector does not forward EVPN routes to peers for which the EVPN SAFI is not enabled, and likewise does not forward IPVPN routes to peers lacking IPVPN SAFI support. For instance, in Figure 8, the Route Reflector does not forward EVPN routes to PE3 if the EVPN SAFI is not enabled on its BGP session with PE3. However, the IPVPN routes are forwarded to all PEs since they all have IPVPN SAFI enabled.

3. IPVPN PE Route Processing

   Regular IPVPN PEs process and import IPVPN routes as specified in [RFC4364] [RFC9252]. For example, PE3 receives only the IPVPN route for prefix IP1/24 and resolves the BGP next-hop to an MPLS/SRv6 tunnel (with IP payload) toward PE1 and/or PE2.

4. Composite PE Route Selection

   Composite PEs MUST perform route selection for prefixes received via multiple ISF SAFIs, applying the procedures described in Section 6:

   * For example, PE4 receives prefix IP1/24 via both an EVPN route and a non-EVPN ISF route (e.g., an IPVPN route). Route selection is performed as specified in Section 6.

   * If the EVPN route is selected, PE4 resolves the BGP next-hop to a tunnel (which may carry either Ethernet or IP payloads) to PE1 and/or PE2. As described in Section 3, the tunnel type used between EVPN PEs depends on the [RFC9136] model supported.

   * Other composite PEs (e.g., PE1 and PE2) receiving the same prefix via both EVPN and IPVPN SAFIs must also apply the route selection process defined in Section 6.

5. Forwarding Behavior Based on Selected Route

   Once a route has been selected for a given IP prefix, packet forwarding MUST follow the forwarding rules associated with the AFI/SAFI of the selected route.

6. Applicability of EVPN Forwarding Enhancements

   In composite domains such as the one depicted in Figure 8, the advanced forwarding features provided by EVPN are available only to composite and EVPN-capable PEs that select an EVPN IP Prefix

route as the best path. These enhancements are not available to IPVPN-only PEs. For example, if PE1 advertises IP1/24 using both EVPN and IPVPN routes, and the EVPN route is selected as the best path, only composite PEs such as PE2 and PE4 can leverage EVPN-specific recursive resolution and forwarding mechanisms [RFC9136]. IPVPN PEs, such as PE3, cannot utilize these capabilities. Consequently, the benefits of EVPN-based indirection and route resolution in large-scale deployments may not be available uniformly across all PEs in the network.

8. Gateway PE Procedures

As defined in Section 3, a gateway PE is an Interworking PE that connects two or more domains and facilitates the re-origination of ISF routes between those domains. Typical examples include data center gateway devices that interconnect domains utilizing different ISF SAFIs, such as EVPN and IPVPN, for the same tenant network.

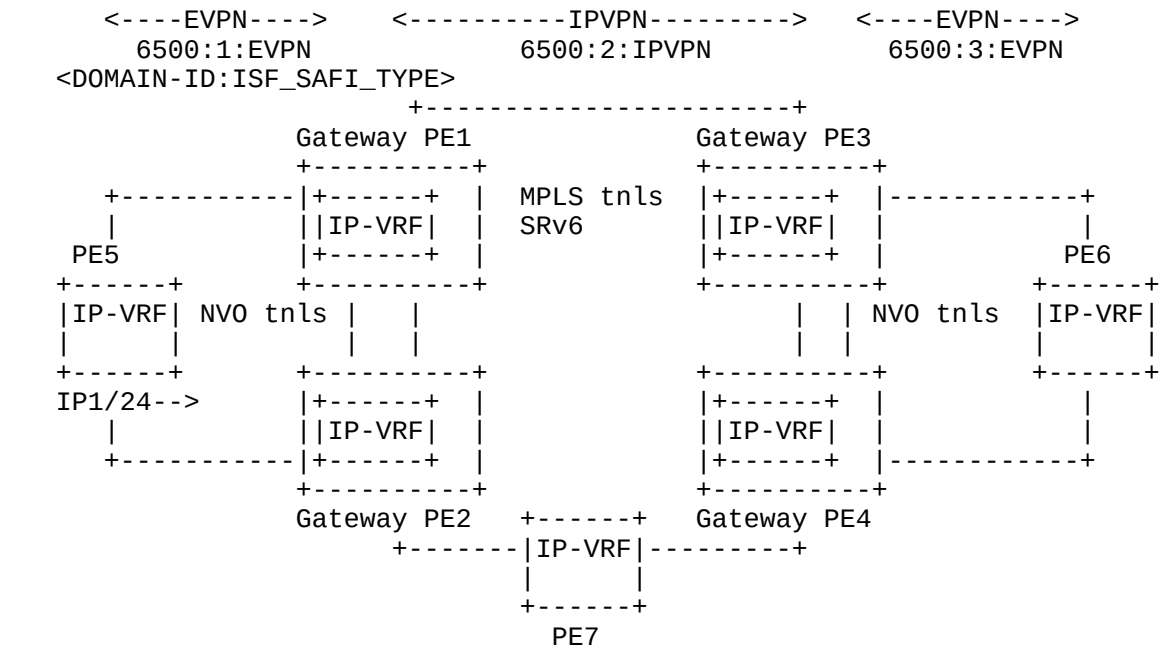
The gateway PE procedures specified in this document define the mechanisms required to support ISF route interconnection across such domains. These procedures extend the concept of a gateway PE beyond the scope of Section 3, which focuses on Layer 2 interconnection, by providing an analogous interconnection model for ISF route exchange at Layer 3.

The procedures described in this section apply to both of the following scenarios:

* Interconnection between domains utilizing different ISF SAFIs (e.g., EVPN to IPVPN).

* Interconnection between domains utilizing the same ISF SAFI (e.g., EVPN to EVPN)

Figure 9 provides an illustrative example of this model, wherein PE1 and PE2 (as well as PE3 and PE4) operate as gateway PEs interconnecting different domains associated with the same tenant.



Note: tnls refer to "tunnels"

Figure 9: Gateway PE example

A gateway PE that is enabled for two ISF SAFIs, referred to here as SAFI x and SAFI y, on the same IP-VRF, MUST follow the procedures described below for re-originating routes between domains.

8.1. Export Conditions

1. A Gateway PE that imports an ISF SAFI x route for prefix P into an IP-VRF MUST export P using ISF SAFI y if all of the following conditions are met:

    a. The route for P is installed in the IP-VRF, indicating that the SAFI x route is well-formed, valid, and selected as the best route.

    b. The PE has an active BGP session with a peer supporting SAFI y, enabled for the same IP-VRF.

    c. Export policy permits the advertisement of the route.

    d. SAFI x and SAFI y are valid ISF SAFIs as defined in Section 3. SAFI x and SAFI y MAY be the same.

Example: In Figure 9, Gateway PEs PE1 and PE2 receive an EVPN IP Prefix route for prefix IP1/24, install the route in their respective IP-VRFs, and re-advertise it using IPVPN.

2. A Gateway PE that receives an ISF SAFI x route for prefix P into an IP-VRF MUST NOT export P using SAFI y under any of the following conditions:

   a. The SAFI x route is not well-formed or valid. Criteria for route validity are defined in the corresponding ISF SAFI specification. For example, an EVPN IP Prefix route that contains both a non-zero ESI and a Gateway IP address is invalid, as specified in [RFC9136], Section 3.2.

   b. The D-PATH attribute of the SAFI x route includes one or more DOMAIN-ID values locally configured on the Gateway PE for the associated IP-VRF. In this case, the route is considered a looped ISF route, as described in Section 4, and MUST NOT be exported using SAFI y.

8.2. Advertisement Behavior

If the export conditions are satisfied, the gateway PE MUST advertise prefix P using ISF SAFI y in accordance with the following procedures:

a. If Uniform Propagation Mode (see Section 5.2) is enabled, the gateway PE MUST follow the procedures defined in Section 5.2, and the gateway PE MUST include the D-PATH attribute when SAFI y is either IPVPN or EVPN. This enables loop detection at downstream gateway PEs.

   When re-originating an ISF route, the gateway PE MUST prepend a <DOMAIN-ID:ISF_SAFI_TYPE> element to the received D-PATH attribute. The DOMAIN-ID reflects the domain from which the route was received, and the ISF_SAFI_TYPE reflects the SAFI of the received route.

   If the received route does not include a D-PATH attribute, the gateway PE MUST create and attach a new D-PATH attribute containing a single segment: the <DOMAIN-ID:ISF_SAFI_TYPE> corresponding to the received route.

Example: In Figure 9, gateway PEs PE1 and PE2 receive an EVPN IP Prefix route from PE5 that does not include a D-PATH attribute. PE1 and PE2 add Domain <6500:1:EVPN> to form the new D-PATH. Gateway PEs PE3 and PE4, upon re-advertising the route, prepend <6500:2:IPVPN>, resulting in PE6 receiving the route with D-PATH {<6500:2:IPVPN>, <6500:1:EVPN>}. This information is then used by PE6 in BGP path selection.

b. The gateway PE uses the Route Distinguisher (RD) of the IP-VRF when re-advertising prefix P via ISF SAFI y.

c. The encapsulation specific context (e.g., label) allocation is a local matter. The gateway PE MAY use per-VRF, per-prefix, or other label allocation models.

d. The gateway PE MUST support the use of distinct Route Target (RT) sets per domain on the same IP-VRF. If multiple domains associated with a tenant use different RT sets, the gateway PE MUST be capable of importing and exporting routes according to each domain’s RT configuration.

e. Although Figure 9 illustrates a scenario with only two domains per gateway PE, gateway PEs may interconnect more than two domains.

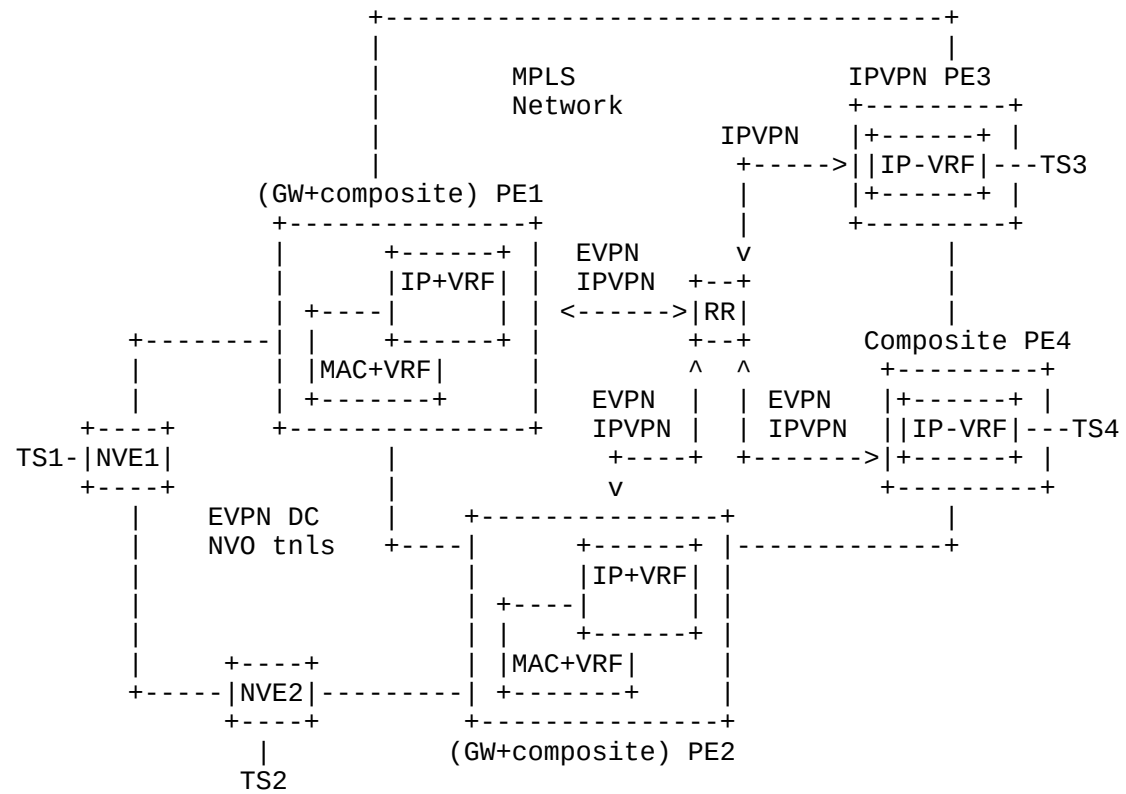
f. There is no restriction on the number of gateway PEs that a given prefix P may traverse before reaching its destination.

g. Informative Note: If prefix P is originated in an EVPN domain and subsequently traverses one or more non-EVPN ISF SAFI domains, it will lose EVPN-specific attributes used for advanced EVPN procedures. For example, if PE1 advertises prefix IP1/24 along with a non-zero ESI (for recursive resolution to that ESI), the ESI value will be reset to zero by the time the route reaches PE6, as it passed through an ISF SAFI domain that is not EVPN-capable. Consequently, certain EVPN-specific functionalities may not be preserved end-to-end.

9. Interworking Use-Cases

While network deployments involving Interworking PEs may align with the scenarios described in Section 7 and Section 8, there are cases where a combination of both gateway PE and composite PE functionality is required. Figure 10 illustrates an example in which gateway PEs also operate as composite PEs. In such scenarios, the devices must not only re-originate ISF routes between domains, such as between EVPN and IPVPN SAFIs or across multiple EVPN domains, but also interoperate with IPVPN-only PEs within domains that include a mix of

composite and IPVPN-only PEs.



Note: tnls refer to "tunnels"

Figure 10: Gateway and composite combined functions - example

In the example illustrated, PE1 and PE2 follow the procedures defined in Section 7 and Section 8. Unlike the scenario described in Section 8, PE1 and PE2 are additionally required to re-originate ISF routes between EVPN domains (i.e., EVPN-to-EVPN), in addition to EVPN-to-IPVPN re-origination. It is important to note that PE1 and PE2 will receive the IP prefix associated with TS4 via both IPVPN and EVPN IP Prefix routes. When re-advertising the selected route to NVE1 and NVE2, PE1 and PE2 MUST apply the D-PATH handling rules and related attribute processing as described in Section 6 (Route Selection Process).

10. BGP Error Handling on Interworking PEs

BGP speakers following this specification MUST adhere to the following error-handling procedures when processing Inter-Subnet Forwarding (ISF) routes:

* Any BGP UPDATE message for an ISF route that includes a D-PATH Path Attribute MUST be handled in accordance with the error-handling rules defined in Section 4 of this document.

* All received BGP UPDATE messages for ISF routes MUST conform to the general error-handling procedures specified in [RFC7606].

* This specification introduces no new error-handling behaviors for BGP UPDATE messages that contain NLRI and BGP Path Attributes defined in other specifications. Implementations SHOULD apply the relevant error-handling rules specified for each supported route type.

If a Gateway PE is configured to propagate BGP Path Attributes for ISF routes between domains, the procedures specified in Section 5.2 are intended to ensure that receiving BGP speakers do not encounter UPDATE messages containing well-formed but semantically inappropriate BGP Path Attributes. However, if a gateway PE incorrectly propagates such attributes in violation of the procedures in Section 5.2, receiving PEs MUST apply the error-handling rules defined in the applicable specifications for the relevant route type and attribute.

The following are examples of such scenarios and their handling:

* If a Gateway PE erroneously propagates the BGP Encapsulation Extended Community or the equivalent Encapsulation TLV in the Tunnel Encapsulation Attribute [RFC9012] from one EVPN domain to another, the receiving PE MAY receive two encapsulation indications with different values. In such a case, the PE MUST follow the procedures in [RFC8365], which permit signaling multiple encapsulation types. As specified in [RFC9012], encapsulations carried via the Tunnel Encapsulation Attribute MUST be treated as equivalent to those conveyed via the Encapsulation Extended Community.

* If a gateway PE propagates an EVPN Extended Community from an EVPN domain into an IPVPN domain, the receiving IPVPN PE MUST ignore such communities, as their semantics are not applicable to the IPVPN SAFI.

* If a gateway PE erroneously propagates a BGP Prefix-SID attribute containing SRv6 Service TLVs [RFC9252] for an ISF route between domains, and the receiving PE receives multiple SRv6 TLV instances, it MUST apply the procedures specified in [RFC9252] for resolving multiple TLVs.

11. Security Considerations

The security considerations outlined in [RFC9136] [RFC8365] for ISF EVPN routes and [RFC4364] for ISF IPVPN routes are applicable to this specification. In addition, the security considerations sections in [RFC9252], [RFC7606], as well as the entire text in [RFC4272] are relevant to this document.

This document introduces the D-PATH Path Attribute (Section 4), which provides a mechanism for control-plane loop prevention when ISF IPVPN and EVPN routes are re-originated across multiple domains via Gateway PEs. When configured and supported correctly, the use of the D-PATH attribute helps prevent both control-plane and data-plane loops. However, incorrect configuration of DOMAIN-ID values or inconsistent support for D-PATH among Gateway PEs may result in false-positive loop detection, traffic discarding, or suboptimal and inconsistent routing behavior. Furthermore, as D-PATH is a transitive BGP attribute, a malicious actor may attempt to inject incorrect domain information that propagates across multiple administrative boundaries.

To mitigate such risks, the use of D-PATH is explicitly restricted to IPVPN and EVPN routes within "walled garden" Virtual Private Networks, as specified in Section 4. A PE that conforms to this specification MUST remove the D-PATH attribute prior to advertising a prefix to a CE router in a SAFI 1 (NLRI used for unicast forwarding) route. If a non-upgraded PE that does not support D-PATH receives such a route and is connected to a CE with Internet access, it may erroneously propagate the D-PATH attribute in a SAFI 1 UPDATE to the CE. If the CE further propagates the route, the D-PATH attribute could inadvertently escape into the public Internet.

However, the presence of the D-PATH attribute in SAFI 1 routes MUST NOT impact BGP best-path selection for those routes and, as such, cannot introduce routing loops or instability in the Internet. Additionally, BGP speakers beyond the "walled garden" that support D-PATH and receive the attribute in SAFI 1 routes MUST apply the "treat-as-withdraw" behavior, as described in Section 4 and consistent with [RFC7606].

As a further safeguard, implementations SHOULD enforce local policy on upgraded PEs to discard any ISF EVPN or IPVPN routes received from non-upgraded peers if such routes include a D-PATH attribute, to prevent unintended propagation. The mechanism by which an implementation determines that ISF EVPN or IPVPN routes are received from non-upgraded peers is outside the scope of this document.

Section 5.2 of this document introduces Uniform Propagation Mode, which enables gateway PEs to propagate a consistent set of BGP Path Attributes across domain boundaries. This mode enhances operational visibility by preserving attributes end-to-end along the route path. However, it also introduces the possibility that an attacker could inject malformed or semantically inappropriate, but syntactically correct, attributes that influence BGP path selection in remote domains.

To mitigate this risk, an operator MAY choose to deploy No-Propagation Mode (Section 5.1), wherein BGP Path Attributes are re-initialized upon domain transition. While this limits attribute-based attack vectors, it also eliminates the ability of downstream PEs to inspect the original set of BGP Path Attributes as intended by the route originator.

Operators SHOULD carefully weigh the trade-offs between visibility and control when selecting the appropriate propagation mode and ensure that policies are in place to validate attribute contents at domain boundaries.

12. IANA Considerations

This document defines a new BGP path attribute known as the BGP Domain Path (D-PATH) attribute.

IANA has assigned a new attribute code type from the "BGP Path Attributes" registry in the "Border Gateway Protocol (BGP) Parameters" registry group:

Path Attribute Value	Code	Reference
36	BGP Domain Path (D-PATH)	[This document]

13. Acknowledgments

The authors want to thank Russell Kelly, Dhananjaya Rao, Suresh Basavarajappa, Mallika Gautam, Senthil Sathappan, Arul Mohan Jovel, Naveen Tubugere, Mathanraj Petchimuthu, Eduard Vasilenko, Amit Kumar, Mohit Kumar, Lukas Krattiger, Gyan Mishra and Stephane Litkowski for their review and suggestions. Thanks to Sue Hares and Jeff Haas as well, for their detailed review to clarify the procedures of the D-PATH attribute. The authors want to also thank especially Gunter van de Velde and Ketan Talaulikar for the thorough review that helped raise the quality of the document significantly.

14. References

14.1. 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>.

[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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[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>.

[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>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <https://www.rfc-editor.org/info/rfc4760>.

[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>.

[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>.

[RFC9252] Dawra, G., Ed., Talaulikar, K., Ed., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)", RFC 9252, DOI 10.17487/RFC9252, July 2022, <https://www.rfc-editor.org/info/rfc9252>.

[RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, DOI 10.17487/RFC4659, September 2006, <https://www.rfc-editor.org/info/rfc4659>.

[RFC7311] Mohapatra, P., Fernando, R., Rosen, E., and J. Uttaro, "The Accumulated IGP Metric Attribute for BGP", RFC 7311, DOI 10.17487/RFC7311, August 2014, <https://www.rfc-editor.org/info/rfc7311>.

14.2. Informative References

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[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>.

[RFC9774] Kumari, W., Sriram, K., Hannachi, L., and J. Haas, "Deprecation of AS_SET and AS_CONFED_SET in BGP", RFC 9774, DOI 10.17487/RFC9774, May 2025, <https://www.rfc-editor.org/info/rfc9774>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <https://www.rfc-editor.org/info/rfc4456>.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <https://www.rfc-editor.org/info/rfc4272>.

[IANA-SAFI] "IANA Subsequent Address Family Identifier Values", <https://www.iana.org/assignments/safi-namespace/safi-namespace.xhtml>.

Authors' Addresses

J. Rabadan (editor)
Nokia
520 Almanor Avenue
Sunnyvale, CA 94085
United States of America
Email: jorge.rabadan@nokia.com

A. Sajassi (editor)
Cisco
225 West Tasman Drive
San Jose, CA 95134
United States of America
Email: sajassi@cisco.com

E. Rosen
Individual
Email: erosen52@gmail.com

J. Drake
Independent
Email: je_drake@yahoo.com

W. Lin
HPE
Email: wen.lin@hpe.com

J. Uttaro
Independent
Email: juttaro@ieee.org

A. Simpson
Nokia
Email: adam.1.simpson@nokia.com