BESS Workgroup
Internet Draft
Intended status: Standards Track

J. Rabadan, Ed.
A. Simpson, Ed.
Nokia

J. Uttaro
AT&T

Expires: September 14, 2017

March 13, 2017

# EVPN Path Attribute Propagation
# draft-rs-bess-evpn-attr-prop-00

Abstract

EVPN is being actively used to provide tenant inter-subnet-forwarding in DC networks, as described in [IP-PREFIX] and [INTER-SUBNET]. When those tenant networks are interconnected to vpn-ipv4/vpn-ipv6 or ipv4/ipv6 BGP networks, there is a need for the EVPN BGP Path Attributes to be seamlessly propagated so that the receiver PE or NVE can consider the original EVPN Attributes in its path calculations. This document analyses the use-cases, requirements and rules based on which the BGP Path Attributes should be propagated between EVPN and other BGP families.

This Internet-Draft will expire on September 14, 2017.



Table of Contents



## 1. Problem Statement

EVPN is being actively used to provide tenant inter-subnet-forwarding in DC networks, as described in [IP-PREFIX] and [INTER-SUBNET]. When those tenant networks are interconnected to vpn-ipv4/vpn-ipv6 or ipv4/ipv6 BGP networks, there is a need for the EVPN BGP Path Attributes to be seamlessly propagated so that the receiver PE or NVE can consider the original EVPN Attributes in its path calculations. This document analyses the use-cases, requirements and rules based on which the BGP Path Attribute propagation should be propagated between EVPN and other BGP families.

EVPN supports the advertisement of ipv4 or ipv6 prefixes in two different route types:

- Route Type 2 - MAC/IP route (only for /32 and /128 host routes), as described by [INTER-SUBNET].

- Route Type 5 - IP Prefix route, as described by [IP-PREFIX].

This proposal describes how the BGP Path Attributes sent along those routes should be propagated to other BGP families being used to advertise tenant IP-Prefixes, such as VPN-IPv4 (AFI/SAFI 1/128), VPN-IPv6 (AFI/SAFI 2/128), IPv4 (AFI/SAFI 1/1) or IPv6 (AFI/SAFI 2/1).

## 2. EVPN Path Attribute Propagation Use Cases

The following Data Center Interconnect (DCI) use-cases have been identified and will be used as a reference in this document.

### 2.1 DCI using a Different Administrative Domain

The assumption in this use-case is that Data Centers (DCs) are connected to other DCs by provider networks that are managed by different administrative entities. While EVPN is used within the DCs to exchange IP Prefixes, the provider interconnect network uses IP-VPN to exchange IP reachability. DC Gateway pairs DGW1 and DGW2 provide a Boundary Router (BR) function between the EVPN and IP-VPN families.

As an example, let's assume NVE1 and NVE2 both advertise an "anycast" prefix A/32. NVE1 uses a Route Type 2 (RT2) or MAC/IP route to encode the A/32 prefix, while NVE1 uses a Route Type 5 (RT5) or IP-Prefix route to encode A/32. DGW1 routers import the routes into the IP-VRF routing table and re-advertise them to the IP-VPN network using a different RD, probably different route-target and their own Next-Hop. DGW2 routers do the opposite translation and re-advertise the host routes using EVPN RT5s. NVE4 uses a PE-CE eBGP session to advertise the host routes to the CE.

While NVEs at DC1 and DC2 set the proper Path Attributes, for example LOCAL_PREFERENCE and Communities 'red' and 'blue', so that NVEs within the DCs can make the right path selection, those Path Attributes are lost when the routes are re-generated at the Boundary Routers (BRs). When the EVPN routes arrive at NVE3 or CE, the path selection cannot be influenced as intended by the NVEs that originated the routes. A set of procedures is needed so that the IP-VPN provider network tunnels all the relevant original EVPN Path Attributes transparently up to the destination EVPN DC.

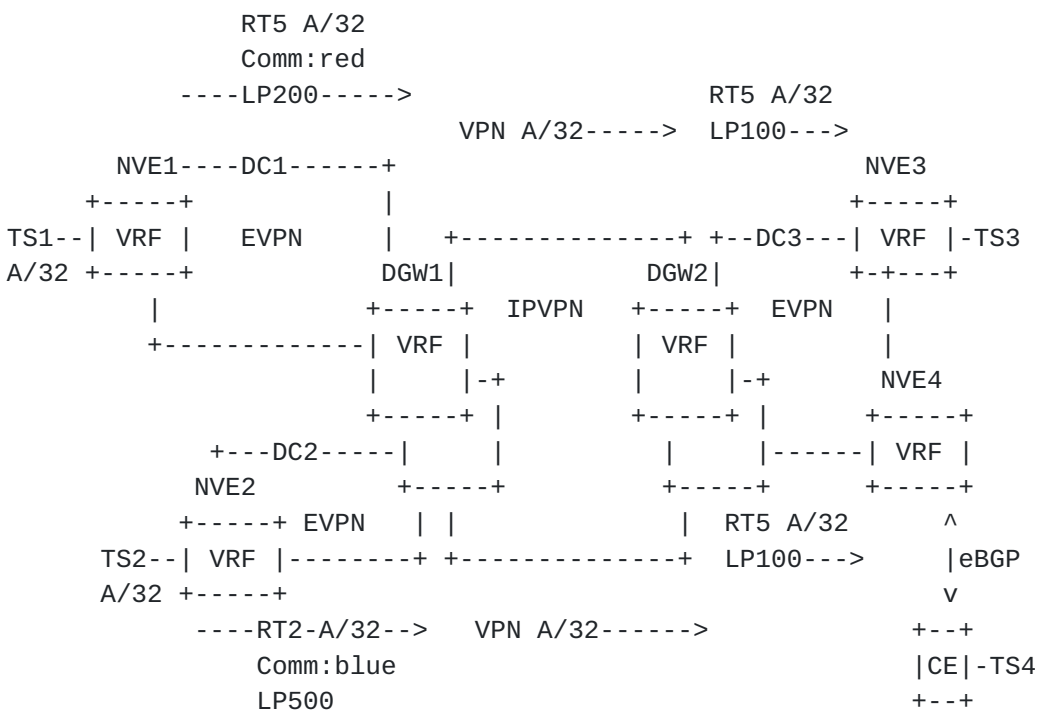


Figure 1 DCI using a Different Administrative Domain

## 2.2 DCI within the Same Administrative Domain

Use-case 2.1 assumed that EVPN DCs were connected using an IP-VPN provider network and there was a need to "tunnel" the original EVPN Path Attributes through the provider IP-VPN network up to the destination EVPN DC. In this section, the entire network is managed by the same entity. The destination PE2 in Figure 2 will receive the two host routes using VPN-IPv4 family directly, even though the routes were originated in the EVPN family.

Multiple models may exist for defining the over-arching VPN solution

defined by this family interaction:

a) In some cases, the BRs (Boundary Routers) need to re-originate the two host routes with the original EVPN Path Attributes (LOCAL_PREFERENCE and Communities in Figure 2) so that they are not lost for PE2's path calculations.

b) In some other cases, the EVPN domains are considered abstracted "CEs" for the IP-VPN network and the BRs just need to reinitialize the Path Attributes so that PE2 does not take the original EVPN Path Attribute into consideration for path calculations.

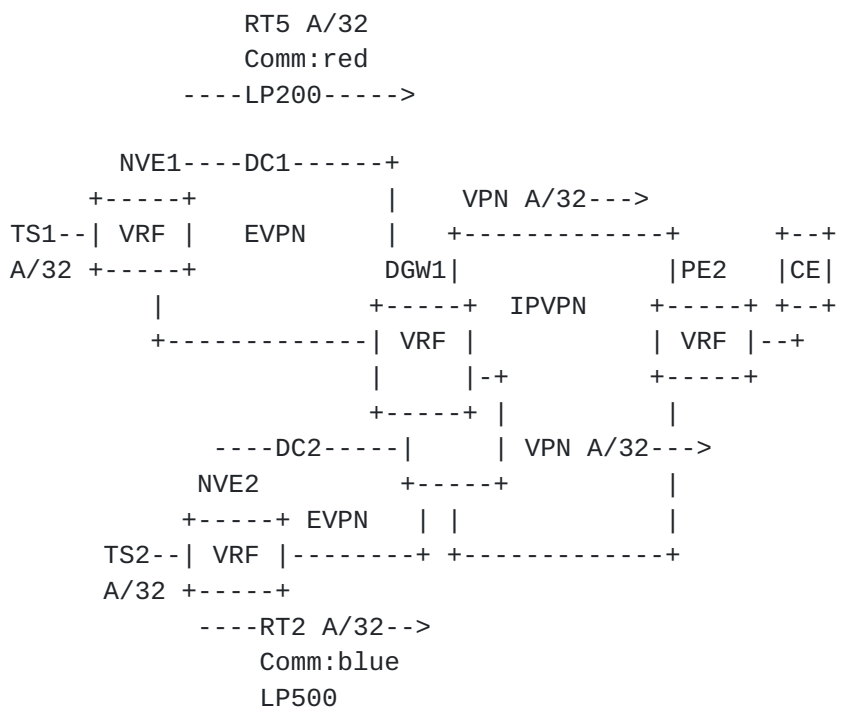


Figure 2 DCI within the Same Administrative Domain

The solution must support both models.

## 2.3 DCI using a Public IP Network

Figure 3 depicts a use-case similar to the one described in section 2.2, however the subnet RT5 is converted to an IPv4 route that gets propagated by BGP throughout a Public Network. As in the previous sections, when the route arrives at the CE router, the originating EVPN Path Attributes are lost. While this may be the desired situation in some cases, in some other cases there is a need to propagate the original EVPN Path Attributes all the way up to the CE router.

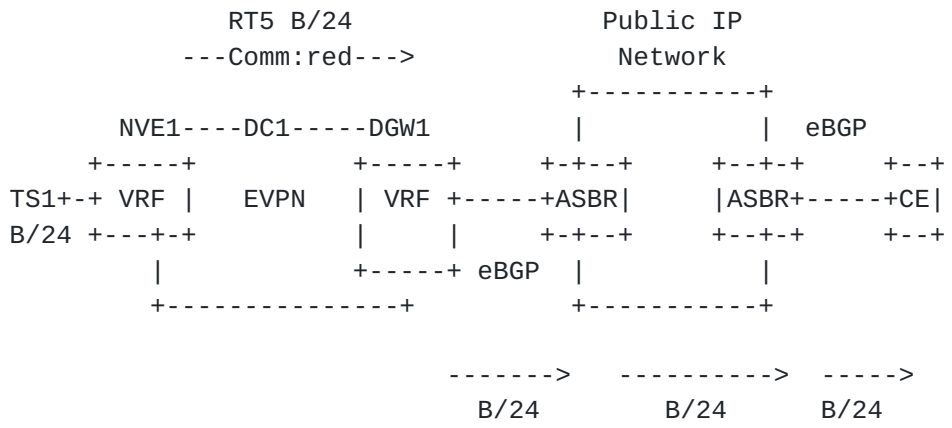


Figure 3 DCI using a Public IP Network

## 3. Solution Requirements

The following requirements have been identified for the Propagation of EVPN Path Attributes:

- The EVPN Path Attribute Propagation solution MUST allow the propagation of path attributes among EVPN (SAFI 70), VPN (SAFI 128) and IP (SAFI 1) families, for IPv4 and IPv6 routes (AFIs 1 and 2).

- The solution SHOULD allow the tunneling of the set of relevant Path Attributes between two BRs of the same family that are connected by another family. Figure 1 provides an example.

- The solution SHOULD allow the propagation of certain key attributes (that are commonly used) between two different families. Figure 2 and 3 show two examples of cases where EVPN Path Attributes should keep accumulating or mapped rather than being tunneled.

## 4. Solution Description

This document proposes three Path Attribute Propagation Modes that satisfy the use-cases and requirements described in sections 2 and 3: No-Propagation-Mode, Tunnel-Mode and Uniform-Mode. In the following sections, the term "BR" or "Boundary Router" refers to the PE router that supports more than one SAFI to manage IP-prefixes in the same IP-VRF and is responsible for the Path Attribute Propagation across families.

### 4.1 EVPN Path Attribute No-Propagation-Mode

This is the default mode of operation. In this mode, the BR will

simply re-initialize the Path Attributes when re-advertising a route to a different SAFI, as though it would for direct or local IP-Prefixes. This model will meet the requirements in those use-cases where the EVPN domain is considered an "abstracted" CE and remote IP-VPN/IP PEs don't need to consider the original EVPN Attributes for path calculations.

### 4.2 EVPN Path Attribute Propagation Tunnel-Mode

In this mode, the Path Attributes are "tunneled" between an ingress and an egress BR. The ingress BR tunnels a set of path attributes for a given family across a provider network that uses a different family. It is typically used for DCs interconnected thru a different administrative domain, as in section 2.1.

The ATT_SET path attribute (defined in RFC6368) is used for this Path Attribute Propagation Tunnel-Mode as follows:

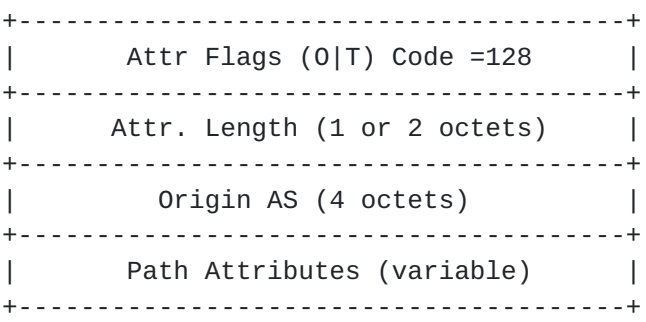


Figure 4 ATT_SET path attribute used for Tunnel-Mode

The following rules MUST be observed:

o These are the Path Attributes that MUST NOT be inserted in the ATT_SET by the ingress BR:
  - MP_REACH_NLRI
  - MP_UNREACH_NLRI
  - NEXT_HOP
  - PTA (PMSI Tunnel Attribute)
  - RFC5512 BGP Encapsulation extended community
  - Tunnel Encapsulation Attribute
  - EVPN-type (0x6) Extended Communities

o ATT_SET insertion rules at ingress BR:

  - IP Prefix routes (RT5 and RT2) learned by the ingress BR on the IP-VRF are imported and re-exported as a different AFI/SAFI with the ATT_SET added.

- The ATT_SET contains an exact copy of all received path attributes except for those that must not be propagated (see bullet above).

- The Origin AS in the attribute encodes the ASN of the exporting VRF.

- Once the ATTR_SET attribute is added to the route, the other path attributes are re-initialized to the basic values that would apply to an exported local/direct IP-VRF route (that is, a route without BGP attributes).

- Note that, compared to RFC6368, in this document ingress BR's IP-VRF does not need IBGP to the CE/NVE. EBGP is possible too. And also, the main focus of this document is EVPN to other families.

o ATT_SET extraction rules at the egress BR:

- The egress BR receiving the ATT_SET, imports the IP-Prefix routes into the IP-VRF, based on the IP-VRF import policies. Different RDs are expected for same routes received from different Next-Hops.

- The Path Attributes in ATT_SET replace the Path Attributes of the received route at the import process (so that the BGP decision process of each IP-VRF considers the original Path Attributes).

- The route, that is re-constructed from ATT_SET, is advertised to the BGP peers of the importing IP-VRF as per [RFC6368]:

  + If the peer is IBGP-based and ATT_SET's Origin AS matches the configured IP-VRF's AS, then the route is advertised "as-is" with Next-Hop-Self (and the original Path Attributes).

  + If the peer is IBGP-based and ATT_SET's Origin AS is different than the configured IP-VRF's AS, then the IBGP-specific Path Attributes are removed, and the ATT_SET Origin AS is prepended to the AS_PATH.

  + If the peer is EBGP-based, then the IBGP-specific Path Attributes are removed and the new AS_PATH will be composed of (ATT_SET Origin AS + received AS_PATH + configured IP-VRF's AS).

### 4.3 EVPN Path Attribute Propagation Uniform-Mode

In this mode, the BR simply keeps accumulating or mapping certain key

commonly used Path Attributes when re-advertising routes to a different family. This mode is typically used for DCs interconnected by the same administrative domain that manages the DCs, as in section 2.2.

The following rules MUST be observed by the BR when propagating Path Attributes:

- The BR imports the routes in the IP-VRF and stores the original Path Attributes. Only the following set of Path Attributes SHOULD be propagated by the BR:
  - AS_PATH
  - IBGP-only Path Attributes: LOCAL_PREF, ORIGINATOR_ID, CLUSTER_ID
  - Communities, (non-EVPN) Extended Communities and Large Communities
- When re-advertising a route to a destination family, the BR MUST copy the AS_PATH of the originating family and prepend the IP-VRF's AS (only for EBGP peers).
- When re-advertising a route to IBGP peers, the BR MUST copy the IBGP-only Path Attributes from the originating family to the re-advertised route.
- Communities, non-EVPN Extended Communities and Large Communities MUST be copied by the BR from the originating family.

Note: the need to include other Path Attributes, such as MED or AIGP, or modify the above behavior will be analyzed in future revisions of this document.

### 4.4 Path Selection across EVPN and IP-VPN

In some cases, an NVE/PE receives the same IP-Prefix from two different families, e.g. EVPN and IP-VPN. This section discusses how the NVE/PE should compare both routes and the rules of selection.

NOTE: this section will be completed in a future revision.

## 5. Deployment Examples

This section will be added in the next revision of the document.

## 6. Conclusions

This document describes the need to propagate EVPN Path Attributes so that NVE/PEs receiving IP-Prefix routes can select paths based on the Attributes that the advertising NVE/PE originally added to the route. In order to achieve that goal, three EVPN Path Attribute Propagation Modes are discussed:

a) No-Propagation-Mode
b) Tunnel-Mode
c) Uniform-Mode

While (a) is the default mode, (b) is required to preserve all the relevant EVPN Path Attributes in use-cases where different Administrative Domains provide connectivity; (c) provides a simple solution to propagate only certain commonly used Path Attributes that are typically used by providers.

This solution will help providers have a seamless EVPN integration in existing IP-VPN and IP networks.

## 6. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

In this document, the characters ">>" preceding an indented line(s) indicates a compliance requirement statement using the key words listed above. This convention aids reviewers in quickly identifying or finding the explicit compliance requirements of this RFC.

## 7. Security Considerations

This section will be added in future versions.

## 8. IANA Considerations

## 9. Terminology

BR: Boundary Router - refers to the router responsible for the Path Attribute Propagation.
RT2: Route Type 2 or MAC/IP route, as per [RFC7432].
RT5: Route Type 5 or IP-Prefix, as per [IP-PREFIX].

## 9. References

### 9.1 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, <http://www.rfc-editor.org/info/rfc7432>.

[RFC6368]Marques, P., Raszuk, R., Patel, K., Kumaki, K., and T. Yamagata, "Internal BGP as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 6368, DOI 10.17487/RFC6368, September 2011, <http://www.rfc-editor.org/info/rfc6368>.

### 9.2 Informative References

[IP-PREFIX] Rabadan et al., "IP Prefix Advertisement in EVPN", draft-ietf-bess-evpn-prefix-advertisement-04, February, 2017.

[INTER-SUBNET] Sajassi et al., "IP Inter-Subnet Forwarding in EVPN", draft-ietf-bess-evpn-inter-subnet-forwarding-03.txt, work in progress, February, 2017

[ENCAP-ATT] Rosen et al., "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-03.txt, work in progress, November, 2016.

## 10. Acknowledgments

## 11. Contributors

## 17. Authors' Addresses

Jorge Rabadan
Nokia
777 E. Middlefield Road
Mountain View, CA 94043 USA
Email: jorge.rabadan@nokia.com

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

Jim Uttaro
AT&T
Email: ju1738@att.com