BESS Working Group S. Litkowski
Internet-Draft S. Agrawal
Intended status: Informational Cisco
Expires: May 7, 2020 K. Patel
Arrcus
S. Zhuang
Huawei
November 4, 2019

Modifying RFC5549 VPNv4 over IPv6 next hop handling procedures
draft-litkowski-bess-vpnv4-ipv6-nh-handling-00

Abstract

RFC4364 and RFC4659 define respectively BGP extensions to provide VPN-IPv4 and VPN-IPv6 services. When defined RFC5549 has brought up an inconsistency in how the next hop is encoded when a VPN-IPv4 NLRI carries an IPv6 next hop compared to RFC4364 and RFC4659. For some reasons, existing and deployed implementations of RFC5549 haven't followed the specification and are using an VPN-IPv6 next hop as in RFC4364 and RFC4659. Moving these implementations to be compliant with RFC5549 may break existing network deployments. This document proposes a modification of RFC5549 to enable compliancy of these implementations. These document also proposes additional modifications of RFC5549 to address missing points.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Table of Contents

1. Problem statement

[RFC4364] and [RFC4659] define respectively BGP extensions to provide VPN-IPv4 and VPN-IPv6 services.

[RFC4364] defines only VPN-IPv4 carried with an IPv4 next hop. For historical reasons, as per Section 4.3.2 of [RFC4364], the next hop address is encoded as a VPN-IPv4 address with an RD of 0. The expected next hop length value is 12 bytes. As stated in Section 4.3.2 of [RFC4364], the justification of using a VPN-IPv4 address in the next hop field came from [RFC2858] which required the next hop address to be in the same address family as the Network Layer Reachability Information.

[RFC4659] defines VPN-IPv6 carried over with either an IPv4 or IPv6 next hop. When IPv4 transport is used, the next hop is encoded as a VPN-IPv6 address with an RD set to 0 followed by the IPv4-mapped IPv6 address of the advertising BGP speaker. The expected next hop length is 24 bytes. When IPv6 transport is used, the next hop is encoded as one or two VPN-IPv6 address(es) still using an RD set to 0. Section 3.2.1.1 of [RFC4659] clearly states: "the BGP speaker SHALL advertise a Next Hop Network Address field containing a VPN-IPv6 address (...) This is potentially followed by another VPN-IPv6 address".

[RFC5549] specifies, among other, the extensions to allow advertising VPN-IPv4 NLRI with an IPv6 protocol next hop. In such a case the next hop of the NLRI is encoded as one or two IPv6 addresses. [RFC5549] does not use VPN-IPv6 addresses but regular IPv6 addresses (no RD) in the next hop field. Refer to Section 4 and Section 6.2 of [RFC5549] for more details. As a consequence, [RFC5549] brings an inconsistency in how the next hop is encoded for VPN SAFIs compared to [RFC4364] and [RFC4659].

While, from a pure specification point of view, this inconsitency between next hop encodings does not create any issue, several existing implementations are using a consistent encoding of the next hop using VPN-IPvx format (using RD set to 0) for all the cases listed above. Authors have looked at nine implementations including the major ones deployed in the market, and all these implementations are encoding the next hop using a VPN-IPv4 format (with RD set to 0), except two which does not support [RFC5549] at all. Although these multiple implementations are not compliant with [RFC5549], modifying these implementations may create backward compatibility issues as well as operation pain for operators who have deployed.

In addition, [RFC5549] only deals with VPN-IPv4 unicast address-family (AFI=1, SAFI=128), but does not handle the case of the VPN-IPv4 multicast address family (AFI=1, SAFI=129).

This document proposes a modification of [RFC5549] for the VPN-IPv4 family to address these problems.

2. Requested modifications

2.1. Modifying next hop encoding

While authors agree that nowadays using a VPN-IP address in a BGP next hop field does not make any sense, to accomodate running codes, deployments and bring consistency with legacy, authors propose [RFC5549] next hop encoding rules to be modified when IPVPN SAFIs are used.

This document proposes to add the following text as part of Section 4 of [RFC5549]:

- o When the AFI=1 and when the SAFI is an IPVPN SAFI (128 or 129), a BGP speaker MUST encode the next hop as VPN-IPv6 address(es) with an RD set to zero.

To accomodate this text, the example provided in Section 6.2 of [RFC5549] must also be modified as follows:

- o Section 6.2 title must be changed to "IPv4 VPN unicast over IPv6 Core"

- o OLD TEXT:

  The MP_REACH_NLRI is encoded with:

  + AFI = 1

  + SAFI = 128

  + Length of Next Hop Network Address = 16 (or 32)

  + Network Address of Next Hop = IPv6 address of Next Hop

  + NLRI = IPv4-VPN routes

- o NEW TEXT:

  The MP_REACH_NLRI is encoded with:

  + AFI = 1

  + SAFI = 128

  + Length of Next Hop Network Address = 24 (or 48)

  + Network Address of Next Hop = IPv6 address of Next Hop encoded as a VPN-IPv6 address with RD set to 0

  + NLRI = IPv4-VPN routes

2.2. Handling of VPN IPv4 multicast SAFI

VPN IPv4 multicast SAFI (AFI=1, SAFI=129) must be handled in the same way as the VPN IPv4 unicast SAFI (AFI=1, SAFI=128).

This document proposes to modify [RFC5549] as follows to accomodate this change:

- Section 3:

  OLD TEXT:

  The following current AFI/SAFI definitions for the IPv4 NLRI or VPN-IPv4 NLRI (<1/1>, <1/2>, <1/4>, and <1/128>) only have provisions for advertising a Next Hop address that belongs to the IPv4 protocol.

  NEW TEXT:

  The following current AFI/SAFI definitions for the IPv4 NLRI or VPN-IPv4 NLRI (<1/1>, <1/2>, <1/4>, <1/128>, and <1/129>) only have provisions for advertising a Next Hop address that belongs to the IPv4 protocol.

- Section 3:

  OLD TEXT:

  This is in addition to the current mode of operation allowing advertisement of NLRI for <AFI/SAFI> of <1/1>, <1/2> and <1/4> with a next hop address of IPv4 type and advertisement of NLRI for <AFI/ SAFI> of <1/128> with a next hop address of VPN-IPv4 type.

  NEW TEXT:

  This is in addition to the current mode of operation allowing advertisement of NLRI for <AFI/SAFI> of <1/1>, <1/2> and <1/4> with a next hop address of IPv4 type and advertisement of NLRI for <AFI/ SAFI> of <1/128> and <1/129> with a next hop address of VPN-IPv4 type.

- Section 3 line "SAFI = 1, 2, 4, or 128" must be changed to "SAFI = 1, 2, 4, 128, or 129".

- Section 4 line "NLRI SAFI = 1, 2, 4, or 128" must be changed to "NLRI SAFI = 1, 2, 4, 128, or 129".

- A Section 6.3 named "IPv4 VPN multicast over IPv6 Core" may be added to provide an example with the following text:

The extensions defined in this document may be used for support of IPv4 VPNs for multicast over an IPv6 backbone. In this application, PE routers would advertise VPN-IPv4 multicast NLRI in the MP_REACH_NLRI along with an IPv6 Next Hop.

The MP_REACH_NLRI is encoded with:

- o AFI = 1
- o SAFI = 129
- o Length of Next Hop Network Address = 24 (or 48)
- o Network Address of Next Hop = IPv6 address of Next Hop encoded as a VPN-IPv6 address with RD set to 0
- o NLRI = IPv4-VPN routes

During BGP Capability Advertisement, the PE routers would include the following fields in the Capabilities Optional Parameter:

- o Capability Code set to "Extended Next Hop Encoding"
- o Capability Value containing <NLRI AFI=1, NLRI SAFI=129, next hop AFI=2>

3. Deployment considerations

As most of the vendors and deployments today are already implementing the VPN-IPv6 address in the next hop field, interoperability in these deployments will not be broken when modifying [RFC5549]. Even if authors have polled multiple vendors including all the major players of the market, there is still a possibility that an existing implementation strictly follows [RFC5549] as it is today. While it should be unlikely, it could happen. In case such a situation exists today, this compliant implementation is not interoperable with most of the implementations of the market and code changes are required (at one side or the other) to get the interoperability. It should be noted that no interoperability issue has been brought to vendors by customers or during interoperability testing between vendors at EANTC for example. By modifying [RFC5549] as we propose, this hypothetical compliant implementation will not be compliant anymore and will require code change to become compliant. This code change can simply be a knob on a per-neighbor basis to accomodate the behavior of the neighbor without breaking any hypothetical deployment between RFC5549 compliant implementations.

As IETF is driven by running code, authors think that changing the existing standard to accomodate running codes and deployments will help the overall industry without causing damages. In case a compliant implementation exists today (but again it is really unlikely), this implementation can add a knob to provide new compliancy and interoperability. This approach will require fewer code changes within the whole industry and then keep most of the existing deployments more stable.

4. Security Considerations

This document does not introduce any additional security issue compared to [RFC4364],[RFC4659] and [RFC5549].

5. Acknowledgements

6. IANA Considerations

IANA has no action.

7. References

7.1. Normative References

[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>.

[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>.

[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>.

[RFC5549] Le Faucheur, F. and E. Rosen, "Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop", RFC 5549, DOI 10.17487/RFC5549, May 2009, <https://www.rfc-editor.org/info/rfc5549>.

[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>.

7.2. Informative References

[RFC2858] Bates, T., Rekhter, Y., Chandra, R., and D. Katz, "Multiprotocol Extensions for BGP-4", RFC 2858, DOI 10.17487/RFC2858, June 2000, <https://www.rfc-editor.org/info/rfc2858>.

Authors' Addresses

Stephane Litkowski
Cisco

Email: slitkows@cisco.com

Swadesh Agrawal
Cisco

Email: swaagrawa@cisco.com

Keyur Patel
Arrcus

Email: keyur@arrcus.com

Shunwan Zhuang
Huawei

Email: zhuangshunwan@huawei.com