S. Litkowski   S. Agrawal   K. Ananthamurthy   K. Patel
Cisco   *Cisco*   *Cisco*   *Arrcus*

RFC 8950
Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop

Abstract

Multiprotocol BGP (MP-BGP) specifies that the set of usable next-hop address families is determined by the Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI). The AFI/SAFI definitions for the IPv4 address family only have provisions for advertising a next-hop address that belongs to the IPv4 protocol when advertising IPv4 Network Layer Reachability Information (NLRI) or VPN-IPv4 NLRI.

This document specifies the extensions necessary to allow the advertising of IPv4 NLRI or VPN-IPv4 NLRI with a next-hop address that belongs to the IPv6 protocol. This comprises an extension of the AFI/SAFI definitions to allow the address of the next hop for IPv4 NLRI or VPN-IPv4 NLRI to also belong to the IPv6 protocol, the encoding of the next hop to determine which of the protocols the address actually belongs to, and a BGP Capability allowing MP-BGP peers to dynamically discover whether they can exchange IPv4 NLRI and VPN-IPv4 NLRI with an IPv6 next hop. This document obsoletes RFC 5549.

Status of This Memo

This is an Internet Standards Track document.

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

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8950.

Copyright Notice

Table of Contents

1. Introduction

Multiprotocol BGP (MP-BGP) [RFC4760] specifies that the set of network-layer protocols to which the address carried in the Next Hop Address field may belong is determined by the Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI). A number of existing AFIs/SAFIs allow the next-hop address to belong to a different address family than the Network Layer Reachability Information (NLRI). For example, the AFI/SAFI <25/65> used (as per [RFC6074]) to perform Layer 2 Virtual Private Network (L2VPN) auto-discovery allows advertising NLRI that contains the identifier of a Virtual Private LAN Service (VPLS) instance or that identifies a particular pool of attachment circuits at a given Provider Edge (PE), while the Next Hop Address field contains the loopback address of a PE. Similarly, the AFI/SAFI <1/132> (defined in [RFC4684]) to advertise Route Target (RT) membership information allows advertising NLRI that contains such RT membership information, while the Next Hop Address field contains the address of the advertising router.

Furthermore, a number of these existing AFIs/SAFIs allow the next hop to belong to either the IPv4 protocol or the IPv6 protocol and specify the encoding of the next-hop information to determine which of the protocols the address actually belongs to. For example, [RFC4684] allows the next-hop address to be either an IPv4 or IPv6 address and states that the Next Hop Address field shall be interpreted as an IPv4 address whenever the length of the next-hop address is 4 octets and as an IPv6 address whenever the length of the next-hop address is 16 octets.

There are situations such as those described in [RFC4925] and [RFC5565] where carriers (or large enterprise networks acting as a carrier for their internal resources) may be required to establish connectivity between 'islands' of networks of one address family type across a transit core of a differing address family type. This includes both the case of IPv6 islands across an IPv4 core and the case of IPv4 islands across an IPv6 core. Where Multiprotocol BGP (MP-BGP) is used to advertise the corresponding reachability information, this translates into the requirement for a BGP speaker to advertise the NLRI of a given address family via a next hop of a different address family (i.e., IPv6 NLRI with an IPv4 next hop and IPv4 NLRI with an IPv6 next hop).

The AFI/SAFI definitions for the IPv6 address family assume that the next-hop address belongs to the IPv6 address family type. Specifically, as per [RFC2545] and [RFC8277], when the <AFI/SAFI> is <2/1>, <2/2>, or <2/4>, the next-hop address is assumed to be of an IPv6 type. As per [RFC4659], when the <AFI/SAFI> is <2/128>, the next-hop address is assumed to be of a VPN-IPv6 type.

However, [RFC4798] and [RFC4659] specify how an IPv4 address can be encoded inside the next-hop IPv6 address field when IPv6 NLRI needs to be advertised with an IPv4 next hop. [RFC4798] defines how the IPv4-mapped IPv6 address format specified in the IPv6 addressing architecture ([RFC4291]) can be used for that purpose when the <AFI/SAFI> is <2/1>, <2/2>, or <2/4>. [RFC4659] defines how the IPv4-mapped IPv6 address format as well as a null Route Distinguisher (RD) can be used for that purpose when the <AFI/SAFI> is <2/128>. Thus, there are existing solutions for the advertisement of IPv6 NLRI with an IPv4 next hop.

Similarly, the AFI/SAFI definitions for the advertisement of IPv4 NLRI or VPN-IPv4 NLRI assume that the next-hop address belongs to the IPv4 address family type. Specifically, as per [RFC4760] and [RFC8277], when the <AFI/SAFI> is <1/1>, <1/2>, or <1/4>, the next-hop address is assumed to be of an IPv4 type. As per [RFC4364], when the <AFI/SAFI> is <1/128>, the next-hop address is assumed to be of a VPN-IPv4 type. As per [RFC6513] and [RFC6514], when the <AFI/SAFI> is <1/129>, the next-hop address is assumed to be of a VPN-IPv4 type. There is clearly no generally applicable method for encoding an IPv6 address inside the IPv4 address field of the next hop. Hence, there is currently no specified solution for advertising IPv4 or VPN-IPv4 NLRI with an IPv6 next hop.

This document specifies the extensions necessary to allow advertisement of IPv4 NLRI or VPN-IPv4 NLRI with a next-hop address that belongs to the IPv6 protocol. This comprises an extension of the AFI/SAFI definitions to allow the address of the next hop for IPv4 NLRI or VPN-IPv4 NLRI to belong to either the IPv4 or the IPv6 protocol, the encoding of the next-hop information to determine which of the protocols the address actually belongs to, and a BGP Capability allowing MP-BGP peers to dynamically discover whether they can exchange IPv4 NLRI and VPN-IPv4 NLRI with an IPv6 next hop. The BGP Capability allows gradual deployment of the functionality of advertising IPv4 reachability via an IPv6 next hop without any flag day nor any risk of traffic black-holing.

This document obsoletes [RFC5549].

1.1. Requirements Language

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

2. Changes Compared to RFC 5549

This document introduces two significant changes compared to [RFC5549]:

- In [RFC5549], when AFI/SAFI <1/128> is used, the next-hop address is encoded as an IPv6 address with a length of 16 or 32 bytes. To accommodate all existing implementations and bring consistency with VPNv4oIPv4 and VPNv6oIPv6, this document modifies how the next-hop address is encoded. The next-hop address is now encoded as a VPN-IPv6 address with a length of 24 or 48 bytes (see Sections 3 and 6.2). This change addresses Erratum ID 5253 ([Err5253]). As all known and deployed implementations are interoperable today and use the new proposed encoding, the change does not break existing interoperability.
- This document allows AFI/SAFI <1/129> (IPv4 multicast) to use an IPv6 underlay using similar encoding and procedures to AFI/SAFI <1/128> (see Sections 3 and 6.3).

3. Extension of AFI/SAFI Definitions for the IPv4 Address Family

As mentioned earlier, MP-BGP specifies that the set of usable next-hop address families is determined by the AFI and the SAFI. The following 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. This document extends the set of usable next-hop address families to include IPv6 in addition to IPv4 when advertising an IPv4 or VPN-IPv4 NLRI.

Specifically, this document allows advertising the MP_REACH_NLRI attribute [RFC4760] with this content:

- AFI = 1
- SAFI = 1, 2, or 4
- Length of Next Hop Address = 16 or 32
- Next Hop Address = IPv6 address of a next hop (potentially followed by the link-local IPv6 address of the next hop). This field is to be constructed as per Section 3 of [RFC2545].
- NLRI = NLRI as per the AFI/SAFI definition

It also allows advertising the MP_REACH_NLRI attribute [RFC4760] with this content:

- AFI = 1
- SAFI = 128 or 129
- Length of Next Hop Address = 24 or 48
- Next Hop Address = VPN-IPv6 address of a next hop with an 8-octet RD set to zero (potentially followed by the link-local VPN-IPv6 address of the next hop with an 8-octet RD set to zero).
- NLRI = NLRI as per the AFI/SAFI definition

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

The BGP speaker receiving the advertisement **MUST** use the Length of Next Hop Address field to determine which network-layer protocol the next-hop address belongs to.

- When the AFI/SAFI is <1/1>, <1/2>, or <1/4> and when the Length of Next Hop Address field is equal to 16 or 32, the next-hop address is of type IPv6.
- When the AFI/SAFI is <1/128> or <1/129> and when the Length of Next Hop Address field is equal to 24 or 48, the next-hop address is of type VPN-IPv6.

Note that this method of using the Length of Next Hop Address field to determine which network-layer protocol the next-hop address belongs to (out of the set of protocols allowed by the AFI/SAFI definition) is the same as that used in [RFC4684] and [RFC6074].

4. Use of BGP Capability Advertisement

[RFC5492] defines a mechanism to allow two BGP speakers to discover if a particular capability is supported by their BGP peer and, thus, whether it can be used with that peer. This document defines a capability that can be advertised using [RFC5492], referred to as the "Extended Next Hop Encoding capability". This capability allows BGP speakers to discover whether, for a given NLRI <AFI/SAFI>, a peer supports advertisement with a next hop whose network protocol is determined by the value of the Length of Next Hop Address field, as specified in Section 3.

A BGP speaker that wishes to advertise an IPv6 next hop for IPv4 NLRI or for VPN-IPv4 NLRI to a BGP peer as per this specification **MUST** use the Capability Advertisement procedures defined in [RFC5492] with the Extended Next Hop Encoding capability to determine whether its peer supports this for the NLRI AFI/SAFI pair(s) of interest. The fields in the Capabilities Optional Parameter **MUST** be set as follows:

- The Capability Code field **MUST** be set to 5 (which indicates the Extended Next Hop Encoding capability).
- The Capability Length field is set to a variable value that is the length of the Capability Value field (which follows).
- The Capability Value field has the following format:

```
        +-----------------------------------------------------+
        | NLRI AFI - 1 (2 octets)                             |
        +-----------------------------------------------------+
        | NLRI SAFI - 1 (2 octets)                            |
        +-----------------------------------------------------+
        | Nexthop AFI - 1 (2 octets)                          |
        +-----------------------------------------------------+
        | .....                                               |
        +-----------------------------------------------------+
        | NLRI AFI - N (2 octets)                             |
        +-----------------------------------------------------+
        | NLRI SAFI - N (2 octets)                            |
        +-----------------------------------------------------+
        | Nexthop AFI - N (2 octets)                          |
        +-----------------------------------------------------+
```

where:

- each triple <NLRI AFI, NLRI SAFI, Nexthop AFI> indicates that the NLRI of <NLRI AFI / NLRI SAFI> may be advertised with a next-hop address belonging to the network-layer protocol of Nexthop AFI.
- the AFI and SAFI values are defined in the "Address Family Numbers" and "Subsequent Address Family Identifier (SAFI) Parameters" registries (see [IANA-AFI] and [IANA-SAFI], respectively).

Since this document only concerns itself with the advertisement of IPv4 NLRI and VPN-IPv4 NLRI with an IPv6 next hop, this specification only allows the following values in the Capability Value field of the Extended Next Hop Encoding capability:

- NLRI AFI = 1 (IPv4)
- NLRI SAFI = 1, 2, 4, 128, or 129
- Nexthop AFI = 2 (IPv6)

This document does not specify the use of the Extended Next Hop Encoding capability with any other combinations of <NLRI AFI, NLRI SAFI, Nexthop AFI>. For example, the Next Hop Encoding capability specified in this document is not intended to be used for NLRI AFIs/SAFIs whose definition already allows use of both IPv4 and IPv6 next hops (e.g., AFI/SAFI = <1/132> as defined in [RFC4684]). Similarly, it is not intended that the Extended Next Hop Encoding capability be used for NLRI AFIs/SAFIs for which there is already a solution for advertising a next hop of a different address family (e.g., AFI/SAFI = <2/1>, <2/2>, or <2/4> with an IPv4 next hop as per [RFC4798] and AFI/SAFI = <2/128> with an IPv4 next hop as per [RFC4659]).

It is expected that if new AFIs/SAFIs are defined in the future, their definitions will have provisions (where appropriate) for both IPv4 and IPv6 next hops from the beginning, with the determination based on the Length of Next Hop Address field. Thus, new AFIs/SAFIs are not expected to make use of the Extended Next Hop Encoding capability.

A BGP speaker **MUST** only advertise the IPv4 or VPN-IPv4 NLRI with an IPv6 next hop to a BGP peer if the BGP speaker has first ascertained via the BGP Capability Advertisement that the BGP peer supports the Extended Next Hop Encoding capability for the relevant AFI/SAFI pair.

The Extended Next Hop Encoding capability provides information about next-hop encoding for a given AFI/SAFI, assuming that AFI/SAFI is allowed. It does not influence whether that AFI/SAFI is indeed allowed. Whether an AFI/SAFI can be used between the BGP peers is purely determined through the Multiprotocol Extensions capability defined in [RFC4760].

5. Operations

By default, if a particular BGP session is running over IPvx (where IPvx is IPv4 or IPv6) and if the BGP speaker sending an update is putting its own address in as the next hop, then the next-hop address **SHOULD** be specified as an IPvx address, using the encoding rules specified in the AFI/SAFI definition of the NLRI being updated. This default behavior may be overridden by policy.

When a next-hop address needs to be passed along unchanged (e.g., as a Route Reflector (RR) would do), its encoding **MUST NOT** be changed. If a particular RR client cannot handle that encoding (as determined by the BGP Capability Advertisement), then the NLRI in question cannot be distributed to that client. For sound routing in certain scenarios, this will require that all the RR clients be able to handle whatever encodings any of them may generate.

6. Usage Examples

6.1. IPv4 over IPv6 Core

The extensions defined in this document may be used as discussed in [RFC5565] for the interconnection of IPv4 islands over an IPv6 backbone. In this application, Address Family Border Routers (AFBRs; as defined in [RFC4925]) advertise IPv4 NLRI in the MP_REACH_NLRI along with an IPv6 next hop.

The MP_REACH_NLRI is encoded with:

- AFI = 1
- SAFI = 1
- Length of Next Hop Address field = 16 (or 32)
- Next Hop Address = IPv6 address of the next hop
- NLRI = IPv4 routes

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

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

6.2. IPv4 VPN Unicast over IPv6 Core

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

The MP_REACH_NLRI is encoded with:

- AFI = 1
- SAFI = 128
- Length of Next Hop Address field = 24 (or 48)
- Next Hop Address = VPN-IPv6 address of a next hop whose RD is set to zero
- NLRI = IPv4-VPN routes

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

- Capability Code set to "Extended Next Hop Encoding"
- Capability Value containing <NLRI AFI=1, NLRI SAFI=128, Nexthop AFI=2>

6.3. IPv4 VPN Multicast over IPv6 Core

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

The MP_REACH_NLRI is encoded with:

- AFI = 1
- SAFI = 129
- Length of Next Hop Address field = 24 (or 48)
- Next Hop Address = VPN-IPv6 address of a next hop whose RD is set to zero
- NLRI = IPv4-VPN routes

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

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

7. IANA Considerations

This document does not define any new code points from those included in [RFC5549].

[RFC5549] added "Extended Next Hop Encoding" to the "Capability Codes" registry ([IANA-CAP-CODE]), which was created by [RFC5492]. IANA has updated the registration of that entry to refer to this document. The value allocated for this Capability Code is 5.

8. Security Considerations

This document does not raise any additional security issues beyond those of BGP-4 and the Multiprotocol Extensions for BGP-4. The same security mechanisms are applicable.

However, as [RFC4272] discusses, BGP is vulnerable to traffic diversion attacks. The ability to advertise an IPv6 next hop adds a new means by which an attacker could cause traffic to be diverted from its normal path. Such an attack differs from preexisting vulnerabilities in that traffic could be forwarded to a distant target across an intervening network infrastructure (e.g., an IPv6 core), allowing an attack to potentially succeed more easily since less infrastructure would have to be subverted. Potential consequences include "hijacking" of traffic or denial of service.

Although not expected to be the typical case, the IPv6 address used as the BGP next-hop address could be an IPv4-mapped IPv6 address (as defined in [RFC4291]). Configuration of the security mechanisms potentially deployed by the network operator (such as security checks on a next-hop address) also need to keep this case in mind.

9. References

9.1. 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>.

[RFC2545]   Marques, P. and F. Dupont, "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545, DOI 10.17487/RFC2545, March 1999, <https://www.rfc-editor.org/info/rfc2545>.

[RFC4291]   Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[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>.

[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>.

[RFC5492]   Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <https://www.rfc-editor.org/info/rfc5492>.

[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>.

[RFC8277]   Rosen, E., "Using BGP to Bind MPLS Labels to Address Prefixes", RFC 8277, DOI 10.17487/RFC8277, October 2017, <https://www.rfc-editor.org/info/rfc8277>.

9.2. Informative References

[Err5253]   RFC Errata, Erratum ID 5253, RFC 5549, <https://www.rfc-editor.org/errata/eid5253>.

[IANA-AFI]   IANA, "Address Family Numbers", <https://www.iana.org/assignments/address-family-numbers/>.

[IANA-CAP-CODE]   IANA, "Capability Codes", <https://www.iana.org/assignments/capability-codes/>.

[IANA-SAFI]   IANA, "Subsequent Address Family Identifiers (SAFI) Parameters", <https://www.iana.org/assignments/safi-namespace/>.

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

[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>.

[RFC4684]   Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard,
            "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol
            Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks
            (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <https://www.rfc-
            editor.org/info/rfc4684>.

[RFC4798]   De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands
            over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", RFC 4798, DOI
            10.17487/RFC4798, February 2007, <https://www.rfc-editor.org/info/rfc4798>.

[RFC4925]   Li, X., Ed., Dawkins, S., Ed., Ward, D., Ed., and A. Durand, Ed., "Softwire Problem
            Statement", RFC 4925, DOI 10.17487/RFC4925, July 2007, <https://www.rfc-
            editor.org/info/rfc4925>.

[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>.

[RFC5565]   Wu, J., Cui, Y., Metz, C., and E. Rosen, "Softwire Mesh Framework", RFC 5565, DOI
            10.17487/RFC5565, June 2009, <https://www.rfc-editor.org/info/rfc5565>.

[RFC6074]   Rosen, E., Davie, B., Radoaca, V., and W. Luo, "Provisioning, Auto-Discovery, and
            Signaling in Layer 2 Virtual Private Networks (L2VPNs)", RFC 6074, DOI
            10.17487/RFC6074, January 2011, <https://www.rfc-editor.org/info/rfc6074>.

[RFC6513]   Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513,
            DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/
            rfc6513>.

[RFC6514]   Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and
            Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/
            RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

Acknowledgments

Authors' Addresses

Stephane Litkowski
Cisco
Email: slitkows@cisco.com

Swadesh Agrawal
Cisco
Email: swaagraw@cisco.com

Krishna Muddenahally Ananthamurthy
Cisco
Email: kriswamy@cisco.com

Keyur Patel
Arrcus
Email: keyur@arrcus.com