Inter-Domain Routing
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2008

F. Le Faucheur
Cisco Systems
E. Rosen
Cisco Systems, Inc.
October 16, 2007

Advertising an IPv4 NLRI with an IPv6 Next Hop
draft-ietf-idr-v4nlri-v6nh-01.txt

Status of this Memo

Copyright Notice

Abstract

MultiProtocol-BGP (MP-BGP) specifies that the set of Network Layer protocols to which the address carried in the Next Hop field may belong is determined by the Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI). The current 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 an IPv4 Network Layer Reachability Information (NLRI) or

a VPN-IPv4 NLRI. This document specifies the extensions necessary to allow advertising an IPv4 NLRI or a 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 an IPv4 NLRI or VPN-IPv4 NLRI to also belong to the IPv6 protocol, the encoding of the Next Hop in order to determine which of the protocols the address actually belongs to, and a new BGP Capability allowing MP-BGP Peers to dynamically discover whether they can exchange IPv4 NLRI and VPN-IPv4 NLRI with an IPv6 Next Hop.

Requirements Language

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].

Table of Contents

1. Introduction

Multi Protocol-BGP (MP-BGP) ([RFC4760]) specifies that the set of Network Layer protocols to which the address carried in the Next Hop field may belong is determined by the Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI). A number of existing AFI/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 [I-D.ietf-l2vpn-signaling] in order to perform L2VPN Autodiscovery allows advertising an NLRI containing the identifier of a VPLS instance, or identifying a particular pool of attachment circuits at a given Provider Edge (PE), while the Next Hop field contains the loopback address of a PE. Similarly, the AFI/SAFI <1/132> defined in [RFC4684] in order to advertise Route Target (RT) membership information, allows advertising an NLRI containing such RT membership information while the Next Hop field contains the address of the advertising router.

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

There are situations such as those described in [RFC4925] and in [I-D.ietf-softwire-mesh-framework] where carriers (or large enterprise networks acting as 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 Multi-Protocol BGP (MP-BGP) is used to advertise the corresponding reachability information, this translates into the requirement for a BGP speaker to advertise Network Layer Reachability Information (NLRI) of a given address family via a Next Hop of a different address family (i.e. IPv6 NLRI with IPv4 Next Hop and IPv4 NLRI with IPv6 Next Hop).

The current 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 [RFC3107], when the <AFI/SAFI> is <2/1>, <2/2>, or <2/4>, the Next Hop address is assumed to be of IPv6 type. As per [RFC4659], when the <AFI/SAFI> is <2/128>, the Next Hop

address is assumed to be of IPv6-VPN type.

However, [RFC4798] and [RFC4659] specify how an IPv4 address can be encoded inside the Next Hop IPv6 address field when an 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 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 current AFI/SAFI definitions for 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 [RFC3107], when the <AFI/SAFI> is <1/1>, <1/2>, or <1/4>, the Next Hop address is assumed to be of IPv4 type. As per [RFC4364], when the <AFI/SAFI> is <1/128>, the Next Hop address is assumed to be of 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 do so. This comprises an extension of the AFI/SAFI definitions to allow the address of the Next Hop for an IPv4 NLRI or VPN-IPv4 NLRI to belong to either the IPv4 or the IPv6 protocol, the encoding of the Next Hop information in order to determine which of the protocols the address

actually belongs to, and a new BGP Capability allowing MP-BGP Peers to dynamically discover whether they can exchange IPv4 NLRI and VPV-IPv4 NLRI with an IPv6 Next Hop. The new BGP capability allows gradual deployment of the new functionality of advertising IPv4 reachability via an IPv6 next hop, without any flag day nor any risk of traffic black-holing.

2. Extension of AFI/SAFI Definitions for IPv4 Address Family

As mentioned earlier, MP-BGP specifies that the set of Network Layer protocols to which the address carried in the Next Hop field may belong is determined by the Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI). 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. This document extends the definition of the AFI/SAFI for advertisement of IPv4 NLRI and VPN-IPv4 NLRI to extend the set of network layer protocols to which

the Next Hop address can belong, to include IPv6 in addition to IPv4.

Specifically, this document allows advertising with [RFC4760] of an MP_REACH_NLRI with:

- AFI=1
- SAFI = 1, 2, 4 or 128
- Length of Next Hop Address = 16 or 32
- Next Hop Address = IPv6 address of 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 current AFI/SAFI definition

This is in addition to the current mode of operation allowing advertisement of an NLRI for <AFI/SAFI> of <1,1>, <1,2> and <1,4> with a next hop address of IPv4 type and advertisement of an NLRI for <AFI/SAFI> of <1,128> with a next hop address of 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 Length of Next Hop Address field is equal to 16 or 32, the next hop address is of type IPv6. Note that this method of using the Length of the 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 used in[RFC4684] and [I-D.ietf-l2vpn-signaling].

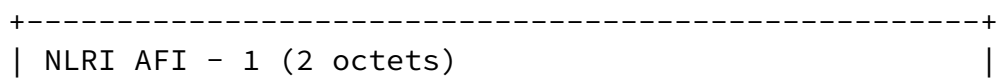
3. Use of BGP Capability Advertisement

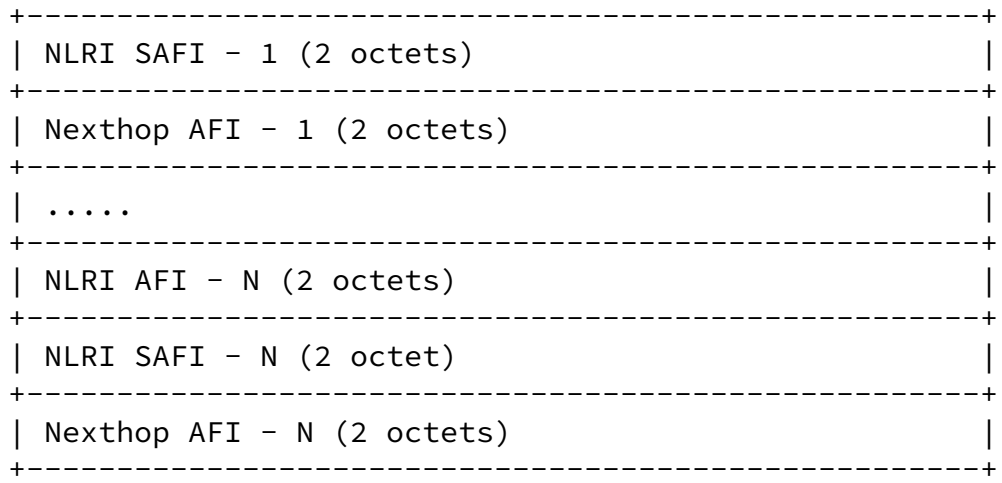
[RFC3392] defines a mechanism to allow two BGP speakers to discover if a particular capability is supported by their BGP peer and thus whether it can be used with that peer. This document defines a new capability which can be advertised using [RFC3392] and which is referred to as the Extended Next Hop 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 2.

A BGP speaker that wishes to advertise to a BGP peer an IPv6 next hop for an IPv4 NLRI or for a VPN-IPv4 NLRI as per this specification MUST use the Capability Advertisement procedures defined in [RFC3392]

with the Extended Next Hop Encoding Capability to establish 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 [To be allocated by IANA] (which indicates the Extended Next Hop Encoding capability).
- The Capability Length field is set to a variable value which is the length of the Capability Value field (which follows).
- The Capability Value field has the following format:





where:

* each triple <NLRI AFI, NLRI SAFI, Nexthop AFI> indicates that an 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 Identifier and Subsequent Address Family Identifier registries maintained by IANA.

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 or 128
- Nexthop AFI=2 (IPv6)

This specification does not propose that the Extended Next Hop Encoding capability be used with any other combinations of NLRI AFI,

NLRI SAFI, Nexthop AFI. In particular, this specification does not propose that the Extended Next Hop Encoding capability be used for NLRI AFI/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 does not propose that the Extended Next Hop Encoding capability be used for NLRI AFI/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 IPv4 next hop as per [RFC4798] and AFI/SAFI=2/128 with IPv4 next hop as per [RFC4659]).

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

A BGP speaker MUST only advertise to a BGP peer an IPv4 or VPN-IPv4 NLRI with an IPv6 Next Hop if the BGP speaker has first ascertained via 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 a AFI/SAFI can be used between the BGP peers is purely determined through the Multiprotocol Extensions capability defined in [RFC4760].

The Extended Next Hop Encoding capability MAY be dynamically updated through the use of the Dynamic Capability capability and associated mechanisms defined in [I-D.ietf-idr-dynamic-cap].

4. 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 (as, e.g., 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.

5. Usage Examples

5.1. IPv4 over IPv6 Core

The extensions defined in this document may be used as discussed in [I-D.ietf-softwire-mesh-framework] for the interconnection of IPV4 islands over an IPv6 backbone. In this application, Address Family Border Routers (AFBR) (as defined in [RFC4925]) advertise IPv4 NLRI information 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 Network Address = 16 (or 32)
- Network Address of Next Hop= IPv6 address of 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>

5.2. IPv4 VPN 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 information 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 Network Address = 16 (or 32)
- Network Address of Next Hop= IPv6 address of Next Hop
- 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. IANA Considerations

This document defines, in section 3, a new Capability Code to indicate the Extended Next Hop Encoding capability in the [RFC3392] Capabilities Optional Parameter. The value for this new Capability Code is to be allocated from the range 1 through 63 set aside for allocation using the "IETF consensus" policy defined in [RFC2434].

7. 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.

8. Acknowledgments

The authors would like to thank Yakov Rekhter, Pranav Metha and John Scudder for their contributions in the approach defined in this document.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2545] Marques, P. and F. Dupont, "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545, March 1999.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[RFC3392] Chandra, R. and J. Scudder, "Capabilities Advertisement with BGP-4", RFC 3392, November 2002.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, February 2006.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

9.2. Informative References

[I-D.ietf-idr-dynamic-cap]
Chen, E. and S. Sangli, "Dynamic Capability for BGP-4", draft-ietf-idr-dynamic-cap-09 (work in progress), November 2006.

[I-D.ietf-l2vpn-signaling]
Rosen, E., "Provisioning, Autodiscovery, and Signaling in L2VPNs", draft-ietf-l2vpn-signaling-08 (work in progress), May 2006.

[I-D.ietf-softwire-mesh-framework]
Wu, J., "Softwire Mesh Framework",

draft-ietf-softwire-mesh-framework-02 (work in progress), July 2007.

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[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, September 2006.

[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, November 2006.

[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, February 2007.

[RFC4925] Li, X., Dawkins, S., Ward, D., and A. Durand, "Softwire Problem Statement", RFC 4925, July 2007.

Authors' Addresses

Francois Le Faucheur
Cisco Systems
Greenside, 400 Avenue de Roumanille
Sophia Antipolis 06410
France

Email: flefauch@cisco.com

Eric Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA 01719
USA

Email: erosen@cisco.com

Full Copyright Statement

Intellectual Property

on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).