

BESS Working Group
Internet-Draft
Intended status: Best Current Practice
Expires: September 23, 2021

G. Mishra
Verizon Inc.
M. Mishra
Cisco Systems
J. Tantsura
L. Wang
Juniper Networks, Inc.
Q. Yang
Arista Networks
A. Simpson
Nokia
S. Chen
Huawei Technologies
March 22, 2021

IPv4 NLRI with IPv6 Next Hop Use Cases
draft-mishra-bess-deplment-guidlin-ipv4nlri-ipv6nh-00

Abstract

As Enterprises and Service Providers upgrade their brown field or green field MPLS/SR core to an IPv6 transport, Multiprotocol BGP (MP-BGP) now plays an important role in the transition of the core as well as edge from IPv4 to IPv6. Operators can now continue to support legacy IPv4, VPN-IPv4, and Multicast VPN-IPv4 customers.

This document describes the critical use case and OPEX savings of being able to leverage the MP-BGP capability exchange usage as a pure transport, allowing both IPv4 and IPv6 to be carried over the same BGP TCP session. By doing so, allows for the elimination of Dual Stacking on the PE-CE connections. Thus making the eBGP peering IPv6-ONLY to now carry both IPv4 and IPv6 Network Layer Reachability Information (NLRI).

This document now provides a solution for IXPs (Internet Exchange points) that are facing IPv4 address depletion at these peering points to use BGP-MP capability exchange defined in [[RFC8950](#)] to carry IPv4 (Network Layer Reachability Information) NLRI in an IPv6 next hop using the [[RFC5565](#)] software mesh framework.

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/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 23, 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
2.	Requirements Language	5
3.	eBGP PE-CE IPv4 and IPv6 NLRI over IPv6 Next Hop Peer Use Case Interop Testing	5
4.	RFC 8950 updates to RFC 5549	6
5.	Operational Improvements with Single IPv6 transport peer . .	7
6.	Operational Considerations	7
7.	IANA Considerations	8
8.	Security Considerations	8
9.	Acknowledgments	8
10.	References	8
10.1.	Normative References	8
10.2.	Informative References	9
Appendix A.	IPv4 NLRI IPv6 Next Hop Vendor Testing	10
A.1.	Router and Switch Vendors Support and Quality Assurance Engineering Lab Results.	11
A.2.	Router and Switch Vendors Interoperability Lab Results. .	11
	Authors' Addresses	11

1. Introduction

As Enterprises and Service Providers upgrade their brown field or green field MPLS/SR core to an IPv6 transport such as MPLS LDPv6, SR-MPLSv6 or SRv6, Multiprotocol BGP (MP-BGP) now plays an important role in the transition of the core from IPv4 to IPv6. Operators can now continue to support legacy IPv4 address family and Sub-Address-Family VPN-IPv4, and Multicast VPN IPv4 customers.

IXPs (Internet Exchange points) are also facing IPv4 address depletion at their peering points, which are large Layer 2 transit backbones that service providers peer and exchange IPv4 and IPv6 (Network Layer Reachability Information) NLRI. Today these transit exchange points are dual stacked. One proposal to solve this issue is to use [[RFC8950](#)] to carry IPv4 (Network Layer Reachability Information) NLRI in an IPv6 next hop and eliminate the IPv4 peering completely using the concept of [[RFC8950](#)] software mesh framework. So now with the MP-BGP reach capability exchanged over IPv4 AFI over IPv6 next hop peer we can now advertise IPv4(Network Layer Reachability Information) NLRI over IPv6 peering using the [[RFC5565](#)] software mesh framework.

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). Historically 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 or VPN-IPv4 Network Layer Reachability Information (NLRI). [[RFC8950](#)] specifies the extensions necessary to allow advertising 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. [[RFC8950](#)] defines the encoding of the Next Hop 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.

With this new MP-BGP capability exchange allows the BGP peering session to act as a pure transport to allow the session to carry Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI) for both IPv4 and IPv6.

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 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 an IPv6 address whenever the length of the Next Hop address is 16 octets.

The current specification for carrying IPv4 Network Layer Reachability Information (NLRI) of a given address family via a Next Hop of a different address family is now defined in [[RFC8950](#)], and 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 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 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.

With the new extensions defined in [[RFC8950](#)] supporting Network Layer Reachability Information (NLRI) and next hop address family mismatch, the BGP peer session can now be treated as a pure transport and carry both IPv4 and IPv6 NLRI at the PE-CE edge over a single IPv6 TCP session. This allows for the elimination of dual stack from the PE-CE peering point, and now allow the peering to be IPv6-ONLY. The elimination of IPv4 on the PE-CE peering points translates into OPEX expenditure savings of point-to-point infrastructure links as well as /31 address space savings and administration and network management of both IPv4 and IPv6 BGP peers. This reduction decreases the number of PE-CE BGP peers by fifty percent, which is a tremendous cost savings for all Enterprises and Service Providers.

While the savings exists at the PE-CE edge, on the core side PE to Route Reflector peering carrying <AFI/SAFI> IPv4 <1/1>, VPN-IPv4 <1/128>, and Multicast VPN <1/129>, the cost savings nets to a break even to be the same as with an IPv4 Core carrying IPv6 NLRI IPv6 <2/1>, VPN-IPv6 <2/128>, and Multicast VPN <2/129>.

This document also provides a possible solution for IXPs (Internet Exchange points) that are facing IPv4 address depletion at these peering points to use BGP-MP capability exchange defined in [[RFC8950](#)] to carry IPv4 (Network Layer Reachability Information) NLRI in an IPv6 next hop using the [[RFC5565](#)] software mesh framework concept of IPv6 NLRI edge over an IPv6 core.

2. 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.

3. eBGP PE-CE IPv4 and IPv6 NLRI over IPv6 Next Hop Peer Use Case Interop Testing

Today the IPv4 NLRI and IPv6 NLRI are carried over separate BGP sessions based on the address family of the NLRI being transported.

The goal of this document is to provide operators interoperability test results from external BGP PE-CE edge peering between vendors Cisco, Juniper, Arista, Nokia and Huawei. The purpose of this document is to prove test data to operators to show that all the features and functionality of carrying IPv4 NLRI over a separate IPv4 peer that exists today is not only viable but recommended to be carried over a single IPv6 peer along with IPv6 NLRI, with no loss of features and functionality using [[RFC8950](#)] IPv6 next hop encoding.

The test results published from this document is to provide concrete evidence that this is now the Best Practice for Edge peering. The defacto standard for operators to now use a single IPv6 peer to carry both IPv4 and IPv6 NLRI.

With the use case defined in this document, IPv6 NLRI Unicast SAFI along with now the IPv4 NLRI Unicast SAFI, can now being carried by the single transport style IPv6 next hop peer.

This document describes the use case of advertising with IPv4 NLRI over IPv6 Next hop with MP_REACH_NLRI with:

- o AFI = 1
- o SAFI = 1
- o Length of Next Hop Address = 16 or 32
- o 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 \[RFC2545\]](#).

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.

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 [\[RFC6074\]](#).

4. [RFC 8950](#) updates to [RFC 5549](#)

This section describes the updates to [\[RFC8950\]](#) next hop encoding from [\[RFC5549\]](#). 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 [\[RFC8950\]](#) (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.

[\[RFC5549\]](#) next hop encoding of MP_REACH_NLRI with:

- o NLRI= NLRI as per current AFI/SAFI definition

Advertising with [\[RFC4760\]](#) MP_REACH_NLRI with:

- o AFI = 1
- o SAFI = 128 or 129
- o Length of Next Hop Address = 16 or 32
- o NLRI= NLRI as per current AFI/SAFI definition

[\[RFC8950\]](#) next hop encoding of MP_REACH_NLRI with:

- o NLRI= NLRI as per current AFI/SAFI definition

Advertising with [\[RFC4760\]](#) MP_REACH_NLRI with:

- o AFI = 1
- o SAFI = 128 or 129
- o Length of Next Hop Address = 24 or 48
- o Next Hop Address = VPN-IPv6 address of 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 is set to zero).

- o NLRI= NLRI as per current AFI/SAFI definition

5. Operational Improvements with Single IPv6 transport peer

As Enterprises and Service Providers migrate their IPv4 core to an MPLS LDPv6 or SRv6 transport, they must continue to be able to support legacy IPv4 customers. With the new extensions defined in [\[RFC4760\]](#), supporting Network Layer Reachability Information (NLRI) and next hop address family mismatch, the BGP peer session can now be treated as a pure transport and carry both IPv4 and IPv6 NLRI at the PE-CE edge. This paves the way to now eliminate dual stacking on all PE-CE peering points to customers making the peering IPv6 only. With this change all IPv4 and IPv6 Network Layer Reachability Information (NLRI) will now be carried over a single BGP session. This also solves the dual stack issue with IXP (Internet Exchange Points) having to maintain separate peering for both IPv4 and IPv6. From an operations perspective the PE-CE edge peering will be drastically simplified with the elimination of IPv4 peers yielding a reduction of peers by 50 percent. From an operations perspective prior to elimination of IPv4 peers an audit is recommended to identify and IPv4 and IPv6 peering incongruencies that may exist and to rectify prior to elimination of the IPv4 peers. No operational impacts or issues are expected with this change.

6. Operational Considerations

With a single IPv6 Peer carrying both IPv4 and IPv6 NLRI there are some operational considerations in terms of what changes and what does not change.

What does not change with a single IPv6 transport peer carrying IPv4 NLRI and IPv6 NLRI below:

Routing Policy configuration is still separate for IPv4 and IPv6 configured by capability as previously

Layer 1, Layer 2 issues such as 1 way fiber or fiber cut will impact both IPv4 and IPv6 as previously.

If the interface is admin down the IPv6 peer would go down and IPv4 NLRI and IPv6 NLRI would be withdrawn as previously.

What does change with a single IPv6 transport peer carrying IPv4 NLRI and IPv6 NLRI below:

Physical interface is no longer dual stacked. Any change in IPv6 address or DAD state will impact both IPv4 and IPv6 NLRI exchange

Single BFD session for both IPv4 and IPv6 NLRI fate sharing as the session is now tied to the transport which now is only IPv6 address family

Both IPv4 and IPv6 peer now exists under the IPv4 address family configuration

Fate sharing of IPv4 and IPv6 address family from a logical perspective now carried over a single IPv6 peer

7. IANA Considerations

There are not any IANA considerations.

8. Security Considerations

The extensions defined in this document allow BGP to propagate reachability information about IPv6 routes over an MPLS IPv4 core network. As such, no new security issues are raised beyond those that already exist in BGP-4 and use of MP-BGP for IPv6. The security features of BGP and corresponding security policy defined in the ISP domain are applicable. For the inter-AS distribution of IPv6 routes according to case (a) of [Section 4](#) of this document, no new security issues are raised beyond those that already exist in the use of eBGP for IPv6 [[RFC2545](#)].

9. Acknowledgments

10. References

10.1. 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>>.

10.2. Informative References

- [I-D.ietf-idr-dynamic-cap]
Ramachandra, S. and E. Chen, "Dynamic Capability for BGP-4", [draft-ietf-idr-dynamic-cap-14](#) (work in progress), December 2011.
- [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., "Software 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, "Software 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>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", [BCP 26](#), [RFC 8126](#), DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [RFC8950] Litkowski, S., Agrawal, S., Ananthamurthy, K., and K. Patel, "Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop", [RFC 8950](#), DOI 10.17487/RFC8950, November 2020, <<https://www.rfc-editor.org/info/rfc8950>>.

[Appendix A](#). IPv4 NLRI IPv6 Next Hop Vendor Testing

IPv4 NLRI with IPv6 Next Hop encoding is supported for all BGP peers both iBGP and eBGP.

This section details the vendor support QA testing of [RFC 8950](#) Next Hop Encoding for "PE-CE eBGP" using GUA (Global Unicast Address), Link Local (LL) peering. This draft's goal is to first ensure that QA

testing of all features and functionality works with "eBGP PE-CE" use case single peer carrying both IPv4 NLRI and IPv6 NLRI and that the routing policy features are all still fully functionality do not change.

A.1. Router and Switch Vendors Support and Quality Assurance Engineering Lab Results.

Vendor	PE-CE eBGP GUI	PE-CE eBGP LL	QA Tested
Cisco	***		
Juniper	***		
Nokia/ALU	***		
Arista	***		
Huawei	***		

Table 1: Vendor Support

A.2. Router and Switch Vendors Interoperability Lab Results.

This section details the vendor interoperability testing and support of [RFC5549](#) that all features and functionality works with "eBGP PE-CE" use case with having a single peer carrying both IPv4 NLRI and IPv6 NLRI and that the routing policy features are fully tested for quality assurance.

Vendor	Cisco	Juniper	Nokia/ALU	Arista	Huawei
Cisco	N/A				
Juniper		N/A			
Nokia/ALU			N/A		
Arista				N/A	
Huawei					N/A

Table 2: Vendor Interop

Authors' Addresses

Gyan Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

Internet-DraDeployment Guidelines for Edge Peering for IPv4- March 2021

Mankamana Mishra
Cisco Systems
821 Alder Drive,
MILPITAS CALIFORNIA 95035

Email: mankamis@cisco.com

Jeff Tantsura
Juniper Networks, Inc.

Email: jefftant.ietf@gmail.com

Lili Wang
Juniper Networks, Inc.
10 Technology Park Drive,
Westford MA 01886
US

Email: liliw@juniper.net

Qing Yang
Arista Networks

Email: qyang@arista.com

Adam Simpson
Nokia

Email: adam.1.simpson@nokia.com

Shuanglong Chen
Huawei Technologies

Email: chenshuanglong@huawei.com

