Interdomain Routing
Internet-Draft
Updates: 2545 (if approved)
Intended status: Standards Track
Expires: 21 February 2026

R. White
Akamai Technologies
J. Tantsura
Nvidia
D. Abraitis
Hostinger

20 August 2025

Link-Local Next Hop Capability for BGP
draft-ietf-idr-linklocal-capability-02

Abstract

To support IPv6 reachability, BGP [RFC4271] relies on the Multiprotocol Extensions as defined in [RFC4760]. [RFC2545] defines the structure of IPv6 next hops. These IPv6 next hops may contain a Global IPv6 address, and optionally can contain an IPv6 Link-Local address when the BGP peer is directly attached and shares a common subnet with the IPv6 Global address.

This document updates [RFC2545] to clarify the encoding of the BGP next hop when the advertising system is directly attached and only an IPv6 Link-Local address is available. A new BGP Capbility [RFC5492] is defined to signal support for this updated encoding.

This clarification applies specifically to IPv6 Link-Local addresses and does not pertain to IPv4 Link-Local addresses as defined in [RFC3927].

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 21 February 2026.

Table of Contents

1. Introduction

To support IPv6 reachability, BGP [RFC4271] relies on the Multiprotocol Extensions as defined in [RFC4760]. [RFC2545] defines the structure of IPv6 next hops. These IPv6 next hops may contain a Global IPv6 address, and optionally can contain an IPv6 Link-Local address when the BGP peer is directly attached and shares a common subnet with the IPv6 Global address.

This document updates [RFC2545] to clarify the encoding of the BGP next hop when the advertising system is directly attached and only an IPv6 Link-Local address is available. A new BGP Capbility [RFC5492] is defined to signal support for this updated encoding.

This clarification applies specifically to IPv6 Link-Local addresses and does not pertain to IPv4 Link-Local addresses as defined in [RFC3927].

BGP speakers are now often deployed on point-to-point links in networks where multihop reachability of any kind is not assumed or desired. (All next hops are assumed to be the reachable through a directly connected point-to-point link.) This is common, for instance, in data center fabrics [RFC7938]. In these situations, a Global IPv6 address is not required for the advertisement of reachability information. In fact, providing Global IPv6 addresses in these kinds of networks can be detrimental to Zero Touch Provisioning (ZTP).

Such BGP deployment models require BGP to run on each link, and any simplification of BGP configuration can simplify orchestration and configuration management. This proposal is a step in that direction.

With this new capability, the need for a Global unicast address assigned to the interfaces is eliminated.

Since IPv6 Link-Local addresses are not required to be globally unique, implementations must ensure that they are strictly associated with a specific interface. (See commentary in [RFC4007].)

2. Link-Local Next Hop Capability

The Link-Local Next Hop capability is a new BGP capability. Its Capability code is 77 and its Capability Length is 0.

A BGP speaker that is willing to use (send and receive) IPv6 Link-Local-only next hops SHOULD advertise the Link-Local Next Hop Capability to its peers only when:

1. It is capable of sending IPv6 Link-Local-only next hops for a route.

2. IPv6 Link-Local neighbors are associated with interfaces as part of their configuration to assist in determining the interface scope of received IPv6 Link-Local-only next hops.

The presence of this capability does not affect the support of Global IPv6 only (16 bytes next hop) and Global IPv6 combined with IPv6 Link-Local (32 bytes next hop), which should continue to be supported as before.

The peers have the flexibility to include both Global and Link-Local BGP IPv6 nexthops, or Link-Local-only IPv6 next hops.

In this document, all procedures described are applicable only when the capability described herein has been successfully advertised by both BGP speakers; i.e., negotiated. When the capability has not been negotiated, the procedures in this document do not apply, and the resulting behavior is considered undefined and out of scope for this specification.

Implementers are encouraged to consult the Appendix for currently known interoperability concerns or incompatibilities when this capability is absent or inconsistently implemented.

3. Changes to the Procedure for Encoding IPv6 Next Hops

Section 2 of [RFC2545] notes IPv6 Link-Local addresses are not generally suitable for use in the Next Hop field of the MP_REACH_NLRI. In order to support the many uses of IPv6 Link-Local addresses, however, [RFC2545] constructs the Next Hop field in IPv6 route advertisements by setting the length of the field to 32, and including both an IPv6 Global and Link-Local address.

[RFC2545] does not, however, provide an explanation for situations where there is only an IPv6 Link-Local address in the Next Hop field of the MP_REACH_NLRI.

If an implementation intends to send a single IPv6 Link-Local forwarding address in the Next Hop field of the MP_REACH_NLRI, it MUST set the length of the Next Hop field to 16 and include only the IPv6 Link-Local address in the Next Hop field.

If an implementation intends to send both a IPv6 Global and Link-Local forwarding address in the Next Hop field of the MP_REACH_NLRI, it MUST set the length of the Next Hop field to 32 and include both the IPv6 Global and Link-Local addresses in the Next Hop field. (XXX Issue-24, do implementations generally have such a knob?) If both the IPv6 Global and Link-Local addresses are carried in the Next Hop Field, the speaker SHOULD provide a local configuration option to determine which address should be preferred for forwarding.

4. IPv6 Next Hop Procedures for Internal and External Peers

Section 5.1.3 of [RFC4271] defines how the NEXT_HOP field is used by BGP for internal and external peering. [RFC2545] does not explicitly discuss next hop procedures in a similar fashion, only the conditions for when the Global IPv6 address is on the same subnet as the peer that a Link-Local IPv6 address is also included in the next hop.

This section defines the behaviors for setting IPv6 next hops when the Link-Local Next Hop Capability has been negotiated between two peers. The next hop MAY consist of only a Link-Local IPv6 next hop.

If, after completing these procedures, there are no IPv6 next hop addresses included in the next hop, the BGP route MUST not be advertised to its peer. Instead, treat-as-withdraw (Section 2 of [RFC7606]) is used.

1. When sending a message to an internal peer, if the route is not locally-originated, the BGP speaker SHOULD NOT modify the Global IPv6 next hop, if one is present, unless it has been explicitly configured to announce its own IP address as the next hop.

   If the internal peer is more than one IP hop away, the BGP speaker MUST NOT include a Link-Local IPv6 next hop.

   If the internal peer is one IP hop away, and the route is not locally-originated, and the route was received from a peer on the same interface as the peer the route is being announced to, the BGP speaker MAY include the received Link-Local IPv6 nexthop for the route. (This is a form of "third-party" next hop.)

   When announcing a locally-originated route to an internal peer, or the BGP speaker has been explicitly configured to announce its own IP address as the next hop, the BGP speaker SHOULD use the Global IPv6 address of the interface of the router through which the announced network is reachable for the speaker as the next hop, if present. If the route is directly connected to the speaker, or if the interface address of the router through which the announced network is reachable for the speaker is the internal peer’s address, the next hop MUST include its own Link-Local IPv6 address.

   If, after evaluating the above procedures, there are no IPv6 next hops included with the route, the route MUST NOT be announced to the remote BGP speaker. (Treat-as-withdraw.)

   These procedures also apply when the BGP speaker is functioning as a route-reflector ([RFC4456]).

2. When sending a message to an external peer, X, and the peer is one IP hop away from the speaker:

   * If the route being announced was learned from an internal peer or is locally-originated, the BGP speaker can use an interface address of the internal peer router (or the internal router) through which the announced network is reachable for the

speaker for the Global IPv6 next hop and shares a common subnet with this address, and/or a Link-Local IPv6 next hop, provided that peer X is directly attached. This is a form of "third party" next hop.

* Otherwise, if the route being announced was learned from an external peer, the speaker can use a Global IPv6 address of any adjacent router (known from the received next hop) that the speaker itself uses for local route calculation in the next hop, provided that peer X shares a common subnet with this Global IPv6 address. Similarly, the speaker can use the received Link-Local IPv6 address, provided that peer X is directly attached. This is a second form of "third party" next hop attribute.

* Otherwise, if the external peer to which the route is being advertised shares a common subnet with one of the interfaces of the announcing BGP speaker, the speaker MAY use the Global IPv6 address associated with such an interface in the next hop. If the external peer is one IP hop away, the announcing BGP speaker SHOULD include a Link-Local IPv6 next hop. These are known as "first party" next hops.

* By default (if none of the above conditions apply), the BGP speaker SHOULD use the IP address of the interface that the speaker uses to establish the BGP connection to peer X in the next hop attribute. The Global IPv6 address, if one is present, SHOULD be included. If the BGP speaker is one IP hop away, the Link-Local IPv6 address SHOULD be included, and MAY be the only next hop address in the next hop. If no next hops are included, the route MUST NOT be announced (treat-as-withdraw).

3. When sending a message to an external peer X, and the peer is multiple IP hops away from the speaker (aka "multihop EBGP"):

   * Link-Local IPv6 next hops MUST NOT be included.

   * The speaker MAY be configured to propagate a Global IPv6 next hop. In this case, when advertising a route that the speaker learned from one of its peers, the Global IPv6 next hop of the advertised route is exactly the same as the next hop attribute of the learned route.

   * By default, the BGP speaker SHOULD use the Global IPv6 address of the interface that the speaker uses in the next hop to establish the BGP connection to peer X.

* If a Global IPv6 next hop is not included, the route MUST NOT be advertised to the external peer (treat-as-withdraw).

5. Error handling

These procedures apply only when the Link-Local Next Hop Capability has been negotiated for a BGP session.

A BGP speaker receiving an MP_REACH_NLRI with the length of the Next Hop Field set to 32, where the update contains anything other than a Global IPv6 followed by a Link-Local IPv6 address, SHOULD do the following.

When both IPv6 next hops contain Link-Local IPv6 addresses, the second Link-Local IPv6 address should be used for forwarding. If both address are not identical, the implementation should bring this to the operator’s attention for debugging.

When the first IPv6 address is 0::0/0 and the second IPv6 address is an IPv6 Link-Local address, the second address is used for forwarding.

If the Next Hop field is malformed, the implementation MUST handle the malformed UPDATE message using the approach of "treat-as-withdraw", as described in section 7.3 of [RFC7606].

If the Next Hop field is properly formed, but the IPv6 Link-Local next hop is not reachable (as determined by an examination of the IPv6 neighbor table), the implementation MAY handle the malformed UPDATE message using the approach of "treat-as-withdraw", as described in section 7.3 of [RFC7606] (see the note above on checking the local neighbor table for the correctness of the next hop).

6. Acknowledgements

The authors would like to thank Vipin Kumar, Dinesh Dutt, Donald Sharp, Jeff Haas, and Brian Carpenter for their contributions to this draft.

This document builds on prior work exploring the use of IPv6 Link-Local addresses as BGP next hops. Notably, [I-D.kumar-idr-link-local-nexthop] and [I-D.kato-bgp-ipv6-link-local] identified operational limitations and proposed mechanisms to enable Link-Local next hop propagation in BGP. These drafts laid the groundwork for defining a standardized capability-based approach, as presented in this document, to ensure interoperable signaling and safe deployment of Link-Local next hops across BGP sessions.

7. IANA Considerations

IANA has assigned capability number 77 for the Link-Local Next Hop Capability described in this document. This registration is in the BGP Capability Codes registry.

Value	Description
77	Link-Local Next Hop Capability

Table 1: Link-Local Next Hop Capability

8. Security Considerations

The mechanism described in this draft can be used as a component of zero-touch provisioning (ZTP) for building BGP peering across point-to-point links. This method, then, can be used by an attacker to form a peering session with a BGP speaker, ultimately advertising incorrect routing information into a routing domain in order to misdirect traffic or cause a denial of service. By using IPv6 Link-Local addresses, the attacker would be able to forgo the use of a valid IPv6 address within the domain, making such an attack easier.

Operators SHOULD carefully consider security when deploying Link-Local addresses for BGP peering. Operators SHOULD filter traffic on links where BGP peering is not intended to occur to prevent speakers from accepting BGP session requests, as well as other mechanisms described in [RFC7454].

Operators MAY also use some form of cryptographic validation on links within the network to prevent unauthorized devices from forming BGP peering sessions. Authentication, such as the TCP authentication [RFC5925], may provide some relief if it is present and correctly configured. However, the distribution and management of keys in an environment where global addresses on BGP speakers are not present may be challenging.

Operators also MAY instruct a BGP peer which has received an UPDATE with an unreachable NEXT_HOP to disable the peering session over which the invalid NEXT_HOP was received pending manual intervention.

9. References

9.1. Normative References

[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>.

[RFC3927] Cheshire, S., Aboba, B., and E. Guttman, "Dynamic Configuration of IPv4 Link-Local Addresses", RFC 3927, DOI 10.17487/RFC3927, May 2005, <https://www.rfc-editor.org/info/rfc3927>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <https://www.rfc-editor.org/info/rfc4456>.

[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>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <https://www.rfc-editor.org/info/rfc5925>.

[RFC7454] Durand, J., Pepelnjak, I., and G. Doering, "BGP Operations and Security", BCP 194, RFC 7454, DOI 10.17487/RFC7454, February 2015, <https://www.rfc-editor.org/info/rfc7454>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[RFC7938] Lapukhov, P., Premji, A., and J. Mitchell, Ed., "Use of BGP for Routing in Large-Scale Data Centers", RFC 7938, DOI 10.17487/RFC7938, August 2016, <https://www.rfc-editor.org/info/rfc7938>.

[RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <https://www.rfc-editor.org/info/rfc7942>.

9.2. Informative References

[I-D.kato-bgp-ipv6-link-local]
明洋, K. and B. Manning, "BGP4+ Peering Using IPv6 Link-local Address", Work in Progress, Internet-Draft, draft-kato-bgp-ipv6-link-local-00, 24 September 2001, <https://datatracker.ietf.org/doc/html/draft-kato-bgp-ipv6-link-local-00>.

[I-D.kumar-idr-link-local-nexthop]
Kumar, V., Mohapatra, P., Dutt, D., and M. Valentine, "BGP Link-Local Next Hop Capability", Work in Progress, Internet-Draft, draft-kumar-idr-link-local-nexthop-02, 13 November 2014, <https://datatracker.ietf.org/doc/html/draft-kumar-idr-link-local-nexthop-02>.

[RFC4007] Deering, S., Haberman, B., Jinmei, T., Nordmark, E., and B. Zill, "IPv6 Scoped Address Architecture", RFC 4007, DOI 10.17487/RFC4007, March 2005, <https://www.rfc-editor.org/info/rfc4007>.

Appendix A. Motiviations for a Capability

Link-Local-only next hops have been inconsistently supported in prior BGP implementations. This capability can permit two conforming implementations to interoperate without additional configuration.

Appendix B. Inconsistency Reports

According to [RFC7942], "This will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature".

FRRouting (https://github.com/frrouting/frr/commit/606fdbb1fab98bac305dca3d19eb38b140b7c3e6) IPv6 next-hop handling when GUA/LL is set to ::/LL.

Bird (https://gitlab.nic.cz/labs/bird/-/commit/17de3a023f7bde293892b41bfafe5740c8553fc8) handling LL/LL case.

Appendix C. Implementation Report

According to [RFC7942], "This will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature".

FRRouting (https://github.com/FRRouting/frr/pull/17871) implementation.

Authors' Addresses

Russ White
Akamai Technologies
Email: russ@riw.us

Jeff Tantsura
Nvidia
Email: jefftant.ietf@gmail.com

Donatas Abraitis
Hostinger
Email: donatas.abraitis@gmail.com