

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: 6 May 2026

B. Decraene, Ed.
Orange
K. Kompella
HPE
S. Krier
Cisco Systems
S. Mohanty
Zscaler
J. G. Scudder, Ed.
K. Wang
HPE
B. Wen
Comcast
2 November 2025

BGP Next Hop Dependent Characteristics Attribute
draft-ietf-idr-nhc-00

Abstract

RFC 5492 allows a BGP speaker to advertise its capabilities to its peer. When a route is propagated beyond the immediate peer, it is useful to allow certain characteristics to be conveyed further. In particular, it is useful to advertise forwarding plane features.

This specification defines a BGP transitive attribute to carry such information, the "Next Hop Dependent Characteristics Attribute," or NHC. Unlike the capabilities defined by RFC 5492, the characteristics conveyed in the NHC apply solely to the routes advertised by the BGP UPDATE that contains the particular NHC.

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 6 May 2026.

Copyright Notice

Copyright (c) 2025 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
1.1. Requirements Language	4
2. BGP Next Hop Dependent Characteristics Attribute	4
2.1. Encoding	4
2.2. Sending the NHC	6
2.2.1. Link-Local-Only Next Hops	7
2.2.2. Aggregation	7
2.3. Receiving the NHC	7
2.4. Attribute Error Handling	8
2.5. Network Operation Considerations	9
3. BGP Identifier Characteristic	9
3.1. Encoding	10
3.2. Sending the BGPID	10
3.2.1. Aggregation	10
3.3. Receiving the BGPID	10
3.3.1. Not Receiving the BGPID	11
3.4. BGPID Error Handling	11
4. IANA Considerations	11
5. Security Considerations	12
5.1. Considerations for the NHC	12
6. References	13
6.1. Normative References	13
6.2. Informative References	14
Appendix A. A Case Where a Link-Local Next Hop Could Lead to a False Positive	15
Acknowledgements	16
Contributors	17
Authors' Addresses	17

1. Introduction

[RFC5492] allows a Border Gateway Protocol (BGP) speaker to advertise its capabilities to its peer. When a route is propagated beyond the immediate peer, it is useful to allow certain characteristics to be conveyed further. In particular, it may be useful to advertise forwarding plane features.

This specification defines a BGP optional transitive attribute to carry such information, the "Next Hop Dependent Characteristics Attribute", or NHC.

Since the NHC is intended chiefly for conveying information about forwarding plane features, it needs to be regenerated whenever the BGP route's next hop is changed. Since, owing to the properties of BGP transitive attributes, this can't be guaranteed (an intermediate router that doesn't implement this specification would be expected to propagate the NHC as opaque data), the NHC encodes the next hop of its originator, or the router that most recently updated the attribute. If the NHC passes through a router that changes the next hop without regenerating the NHC, they will fail to match when later examined, and the recipient can act accordingly. This scheme allows NHC support to be introduced into a network incrementally. Informally, the intent is that,

- * If a router is not changing the next hop, it can obviously propagate the NHC just like any other optional transitive attribute.
- * If a router is changing the next hop, then it has to be able to vouch for every characteristic it includes in the NHC.

Complete details are provided in Section 2.

An NHC carried in a given BGP UPDATE message conveys information that relates to all Network Layer Reachability Information (NLRI) advertised in that particular UPDATE, and only to those NLRI. A different UPDATE message originated by the same source might not include an NHC, and if so, the NLRI carried in that UPDATE would not be affected by the NHC. By implication, if a router wishes to use NHC to describe all NLRI it originates, it needs to include an NHC with each UPDATE it sends.

Informally, a characteristic included in a given NHC should not be thought of as a characteristic of the next hop, but rather a characteristic of the path, which depends on the ability of the next hop to support it. Hence, it is said to be "dependent on" the next hop.

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. BGP Next Hop Dependent Characteristics Attribute

2.1. Encoding

The BGP Next Hop Dependent Characteristics attribute (NHC attribute, or just NHC) is an optional, transitive BGP path attribute with type code 39. The NHC always includes a network layer address identifying the next hop of the route the NHC accompanies. The NHC signals potentially useful information related to the forwarding plane features, so it is desirable to make it transitive to ensure propagation across BGP speakers (e.g., route reflectors) that do not change the next hop and are therefore not in the forwarding path. The next hop data is to ensure correctness if it traverses BGP speakers that do not understand the NHC. This is further explained below.

The Attribute Data field of the NHC attribute is encoded as a header portion that identifies the router that created or most recently updated the attribute, followed by one or more Type-Length-Value (TLV) triples:

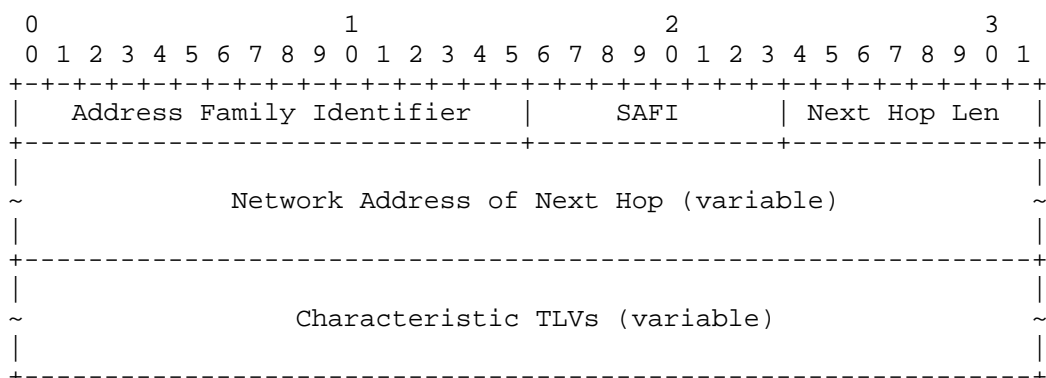


Figure 1: NHC Format

The meanings of the header fields (Address Family Identifier, SAFI or Subsequent Address Family Identifier, Length of Next Hop, and Network Address of Next Hop) are as given in Section 3 of [RFC4760].

In turn, each Characteristic is a TLV:

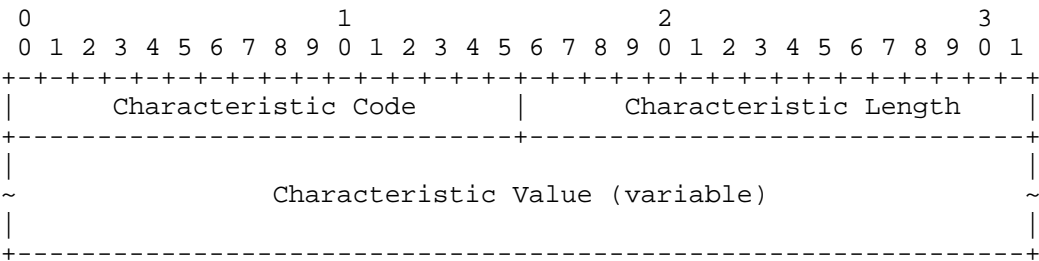


Figure 2: Characteristic TLV Format

Characteristic Code: a two-octet unsigned integer that indicates the type of characteristic advertised and unambiguously identifies an individual characteristic.

Characteristic Length: a two-octet unsigned integer that indicates the length, in octets, of the Characteristic Value field. A length of 0 indicates that the Characteristic Value field is zero-length, i.e., it has a null value.

Characteristic Value: a variable-length field. It is interpreted according to the value of the Characteristic Code.

A BGP speaker MUST NOT include more than one instance of a characteristic with the same Characteristic Code, Characteristic Length, and Characteristic Value. Note, however, that processing multiple instances of such a characteristic does not require special handling, as additional instances do not change the meaning of the announced characteristic; thus, a BGP speaker MUST be prepared to accept such multiple instances.

BGP speakers MAY include more than one instance of a characteristic (as identified by the Characteristic Code) with different Characteristic Values. Processing of these characteristic instances is specific to the Characteristic Code and MUST be described in the document introducing the new characteristic.

Characteristic TLVs MUST be placed in the NHC in increasing order of Characteristic Code. (In the event of multiple instances of a characteristic with the same Characteristic Code as discussed above, no further sorting order is defined here.) Although the major sorting order is mandated, an implementation MUST be prepared to consume characteristics in any order, for robustness reasons.

2.2. Sending the NHC

Suppose a BGP speaker S has a route R it wishes to advertise with next hop N to its peer.

If S is originating R into BGP, it MAY include an NHC attribute with it, that carries characteristic TLVs that describe aspects of R. S MUST set the next hop depicted in the header portion of the NHC to be equal to N, using the encoding given above.

If S has received R from some other BGP speaker, two possibilities exist. First, S could be propagating R without changing N. In that case, S does not need to take any special action; it SHOULD simply propagate the NHC unchanged unless specifically configured otherwise. Indeed, we observe that this is no different from the default action a BGP speaker takes with an unrecognized optional transitive attribute -- it is treated as opaque data and propagated.

Second, S could be changing R in some way, and in particular, it could be changing N. If S has changed N, it MUST NOT propagate the NHC unchanged. It SHOULD include a newly-constructed NHC attribute with R, constructed as described above in the "originating R into BGP" case. Any given characteristic TLV carried by the newly-constructed NHC attribute might use information from the received NHC attribute as input to its construction, possibly as straightforwardly as simply copying the TLV. The details of how the characteristics in the new NHC are constructed are specific to the definition of each characteristic. Any characteristic TLVs received by S that are for characteristics not supported by S will not be included in the newly-constructed NHC attribute S includes with R.

An implementation SHOULD propagate the NHC and its contained characteristics by default. An implementation SHOULD provide configuration control of whether any given characteristic is propagated. An implementation MAY provide finer-grained control on propagation based on attributes of the peering session, as discussed in Section 5.1.

Due to the nature of BGP optional transitive path attributes, any BGP speaker that does not implement this specification will propagate the NHC, the requirements of this section notwithstanding. Such a speaker will not update the NHC, however.

Certain NLRI formats do not include a next hop at all, one example being the Flow Specification NLRI [RFC8955]. The NHC MUST NOT be sent with such NLRI.

2.2.1. Link-Local-Only Next Hops

In some cases, the BGP speaker sending a route might encode only a link-local address and no global address. In such a case, a problem arises because there is no expectation of global uniqueness of such an address, and the "semantic match" discussed in Section 2.3 could yield a false positive. An illustration is provided in Appendix A.

To mitigate this problem, if a BGP speaker originates a route whose next hop has no global part, it **MUST** include a BGPID TLV (Section 3).

2.2.2. Aggregation

When aggregating routes, the above rules for constructing a new NHC **MUST** be followed. The decision of whether to include the NHC with the aggregate route and what its form will be depends in turn on whether any characteristics are eligible to be included with the aggregate route. If there are no eligible characteristics, the NHC **MUST NOT** be included.

The specification for an individual characteristic must define how that characteristic is to be aggregated. If no rules are defined for a given characteristic, that characteristic **MUST NOT** be aggregated.

(Route aggregation is described in [RFC4271]. Although prefix aggregation -- combining two or more more-specific prefixes to form one less-specific prefix -- is one application of aggregation, we note that another is when two or more routes for the same prefix are selected to be used for multipath forwarding.)

2.3. Receiving the NHC

An implementation receiving routes with an NHC **SHOULD NOT** discard the attribute or its contained characteristics by default. An implementation **SHOULD** provide configuration control of whether any given characteristic is processed. An implementation **MAY** provide finer-grained control on propagation based on attributes of the peering session, as discussed in Section 5.1.

When a BGP speaker receives a BGP route that includes the NHC, it **MUST** compare the address given in the header portion of the NHC and illustrated in Figure 1 to the next hop of the BGP route. If the two match, the NHC may be further processed. If the two do not match, it means that some intermediate BGP speaker that handled the route in transit both does not support NHC and changed the next hop of the route. In this case, the contents of the NHC cannot be used, and the NHC **MUST** be discarded without further processing, except that the contents **MAY** be logged.

In considering whether the next hop "matches", a semantic match is sought. While bit-for-bit equality is a trivial test of matching, there may be certain cases where the two are not bit-for-bit equal, but still "match". An example is when an MP_REACH Next Hop encodes both a global and a link-local IPv6 address. In that case, the link-local address might be removed during Internal BGP (IBGP) propagation, but the two would still be considered to match if they were equal on the global part. See Section 3 of [RFC2545]. In other cases, only a link-local address might be present. This is discussed in Section 2.2.1; in such a case, further information is required to permit matching. This is discussed in Section 3.

A BGP speaker receiving a Characteristic Code that it supports behaves as defined in the document defining the Characteristic Code. A BGP speaker receiving a Characteristic Code that it does not support MUST ignore that Characteristic Code. In particular, the receipt of an unrecognized Characteristic Code MUST NOT be handled as an error.

The presence of a characteristic SHOULD NOT influence route selection or route preference, unless tunneling is used to reach the BGP next hop, the selected route has been learned from External BGP (that is, the next hop is in a different Autonomous System), or by configuration (see following). Indeed, it is in general impossible for a node to know that all BGP routers of the Autonomous System (AS) will understand a given characteristic, and if different routers within an AS were to use a different preference for a route, forwarding loops could result unless tunneling is used to reach the BGP next hop. Following this reasoning, if the administrator of the network is confident that all routers within the AS will interpret the presence of the characteristic in the same way, they could relax this restriction by configuration.

2.4. Attribute Error Handling

An NHC is considered malformed if the length of the attribute, encoded in the Attribute Length field of the BGP Path Attribute header (Section 4.3 of [RFC4271]), is inconsistent with the lengths of the contained characteristic TLVs. In other words, the sum of the sizes (Characteristic Length plus 4) of the contained characteristic TLVs, plus the length of the NHC header (Figure 1), must be equal to the overall Attribute Length.

A BGP UPDATE message with a malformed NHC SHALL be handled using the approach of "attribute discard" defined in [RFC7606].

Unknown Characteristic Codes MUST NOT be considered to be an error.

An NHC that contains no characteristic TLVs MAY be considered malformed, although it is observed that the prescribed behavior of "attribute discard" is semantically no different from that of having no TLVs to process. There is no reason to propagate an NHC that contains no characteristic TLVs.

A document that specifies a new NHC Characteristic should provide specifics regarding what constitutes an error for that NHC Characteristic.

If a characteristic TLV is malformed, that characteristic TLV SHOULD be ignored and removed. Other characteristic TLVs SHOULD be processed as usual. If a given characteristic TLV requires different error-handling treatment than described in the previous sentences, its specification should provide specifics.

2.5. Network Operation Considerations

In the corner case where multiple nodes use the same IP address as their BGP next hop, such as with anycast nodes as described in [RFC4786], a BGP speaker MUST NOT advertise a given characteristic unless all nodes sharing this same IP address support this characteristic. The network operator operating those anycast nodes is responsible for ensuring that an anycast node does not advertise a characteristic not supported by all nodes sharing this anycast address. The means for accomplishing this are beyond the scope of this document.

In cases where a BGP speaker receives a route for some prefix P with next hop N that carries an NHC, and receives a different route for P, N that carries no NHC or a NHC with conflicting content, that could be indicative of a configuration error as described above. In such a case, an implementation MAY log an error to help diagnose the potential problem.

3. BGP Identifier Characteristic

As discussed in Section 2.2.1, it might be possible that a route could be originated that has no global part in its next hop. To provide uniqueness in this case, it is sufficient to associate the BGP Identifier and AS Number of the route's sender. The BGP Identifier Characteristic (BGPID) provides a way to convey this information if required.

3.1. Encoding

The BGPID has characteristic code 3, characteristic length 8, and carries as its value the BGP Identifier and Autonomous System Number of its sender:

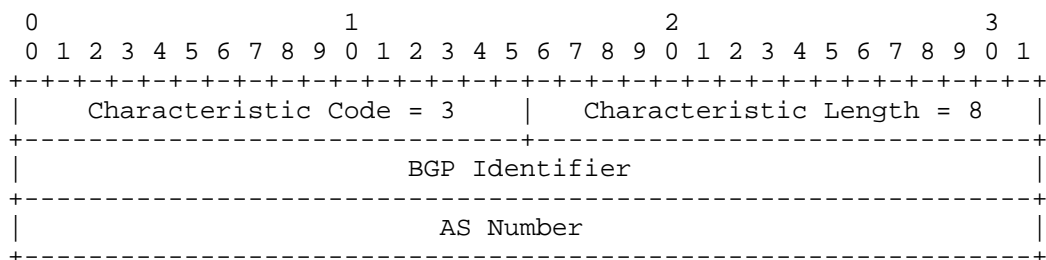


Figure 3: BGPID TLV Format

BGP Identifier: The BGP Identifier (Section 4.2 of [RFC4271], and [RFC6286]) of the route's sender.

AS Number: The Autonomous System Number [RFC6793] of the route's sender. In cases where the sender might represent different Autonomous System Numbers to different peers (for example, [RFC5065], [RFC7705]), the value used is the one that was in the sender's BGP OPEN to the peer concerned.

3.2. Sending the BGPID

Under the circumstances described in Section 2.2.1, the BGPID MUST be included. Under other circumstances, the BGPID MAY be included.

3.2.1. Aggregation

Since the BGPID, by definition, is regenerated whenever the next hop is changed and provides context to disambiguate the next hop carried in the NHC header, there is no case in which it might need to be aggregated.

3.3. Receiving the BGPID

Under the circumstances described in Section 2.2.1, a NEXT_HOP received from a given peer MUST NOT be considered a "semantic match" for the NHC unless the BGP Identifier and Autonomous System of that peer match the BGP Identifier and Autonomous System carried in the BGPID.

Since the only case in which the BGPID might be needed to disambiguate the next hop carried in the NHC involves the immediate peer (see Appendix A for more detail), the BGP Identifier and Autonomous System of the peer are readily derived; they are the values that were received in that peer’s OPEN message.

Other uses of the BGPID are beyond the scope of this document. In particular, if a route is received that has a global part to its NEXT_HOP and thus does not match the circumstances described in Section 2.2.1, but which nonetheless has a BGPID, this specification requires no specific action. In such a case, the BGPID can be disregarded.

3.3.1. Not Receiving the BGPID

Under the circumstances described in Section 2.2.1, if a BGPID is not present in the NHC, the next hop match described in Section 2.3 MUST be considered to have failed.

3.4. BGPID Error Handling

The BGPID is considered malformed and must be disregarded if its length is other than eight.

If more than one instance of the BGPID is included in an NHC, instances beyond the first MUST be disregarded.

The situation where a route is received that fails the test described in Section 3.3 is not an error. However, it might indicate a misconfiguration in the network, and a message MAY be logged.

4. IANA Considerations

IANA has made a temporary allocation in the BGP Path Attributes registry of the Border Gateway Protocol (BGP) Parameters group. IANA is requested to make this allocation permanent and to update its name and reference as shown below.

Value	Code	Reference
39	BGP Next Hop Dependent Characteristic (NHC)	(this doc)

Table 1

IANA is requested to create a new registry called "BGP Next Hop Dependent Characteristic Codes" within the Border Gateway Protocol (BGP) Parameters group. The registry's allocation policy is First Come, First Served, except where designated otherwise in Table 2. It is seeded with the following values:

Value	Description	Reference	Change Controller
0	reserved	(this doc)	IETF
1	ELCv3	draft-ietf-idr-elc-00	IETF
2	NNHN	draft-wang-idr-next-next-hop-nodes-01	kfwang@juniper.net
3	BGPID	(this doc)	IETF
4	IFIT	draft-ietf-idr-bgp-ifit-capabilities-05	IETF
5	AMetric	draft-ietf-idr-bgp-generic-metric-01	IETF
65400 - 65499	private use	(this doc)	IETF
65500 - 65534	reserved for experimental use	(this doc)	IETF
65535	reserved	(this doc)	IETF

Table 2

5. Security Considerations

5.1. Considerations for the NHC

The header portion of the NHC contains the next hop the attribute's originator included when sending it, or that an intermediate router included when updating the attribute (in the latter case, the "contract" with the intermediate router is that it performed the checks in Section 2.3 before propagating the attribute). This will typically be an IP address of the router in question. This may be an

infrastructure address the network operator does not intend to announce beyond the border of its Autonomous System, and it may even be considered in some weak sense confidential information.

A motivating application for this attribute is to convey information between Autonomous Systems that are under the control of the same administrator. In such a case, it would not need to be sent to other Autonomous Systems. At the time of writing, work [I-D.uttaro-idr-bgp-oad] is underway to standardize a method of distinguishing between the two categories of external Autonomous Systems, and if such a distinction is available, an implementation can take advantage of it by constraining the NHC and its contained characteristic to only propagate by default to and from the former category of Autonomous Systems. If such a distinction is not available, a network operator may prefer to configure routers peering with Autonomous Systems not under their administrative control to not send or accept the NHC or its contained characteristic, unless there is an identified need to do so.

The foregoing notwithstanding, control of NHC propagation can't be guaranteed in all cases -- if a border router doesn't implement this specification, the attribute, like all BGP optional transitive attributes, will propagate to neighboring Autonomous Systems. (This can be seen as a specific case of the general "attribute escape" phenomenon discussed in [I-D.haas-idr-bgp-attribute-escape].) Similarly, if a border router receiving the attribute from an external Autonomous System doesn't implement this specification, it will store and propagate the attribute, the requirements of Section 2.3 notwithstanding. So, sometimes this information could leak beyond its intended scope. (Note that it will only propagate as far as the first router that does support this specification, at which point it will typically be discarded due to a non-matching next hop, per Section 2.3.)

If the attribute leaks beyond its intended scope, characteristics within it would potentially be exposed. Specifications for individual characteristics should consider the consequences of such unintended exposure, and should identify any necessary constraints on propagation.

6. References

6.1. 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/rfc/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/rfc/rfc2545>>.
- [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/rfc/rfc4271>>.
- [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/rfc/rfc4760>>.
- [RFC6286] Chen, E. and J. Yuan, "Autonomous-System-Wide Unique BGP Identifier for BGP-4", RFC 6286, DOI 10.17487/RFC6286, June 2011, <<https://www.rfc-editor.org/rfc/rfc6286>>.
- [RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", RFC 6793, DOI 10.17487/RFC6793, December 2012, <<https://www.rfc-editor.org/rfc/rfc6793>>.
- [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/rfc/rfc7606>>.
- [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/rfc/rfc8174>>.

6.2. Informative References

- [I-D.haas-idr-bgp-attribute-escape]
Haas, J., "BGP Attribute Escape", Work in Progress, Internet-Draft, draft-haas-idr-bgp-attribute-escape-03, 9 April 2025, <<https://datatracker.ietf.org/doc/html/draft-haas-idr-bgp-attribute-escape-03>>.
- [I-D.ietf-idr-next-hop-capability]
Decraene, B., Kompella, K., and W. Henderickx, "BGP Next-Hop dependent capabilities", Work in Progress, Internet-Draft, draft-ietf-idr-next-hop-capability-08, 8 June 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-idr-next-hop-capability-08>>.

[I-D.scudder-bgp-entropy-label]

Scudder, J. and K. Kompella, "BGP Entropy Label Capability, Version 2", Work in Progress, Internet-Draft, draft-scudder-bgp-entropy-label-00, 28 April 2022, <<https://datatracker.ietf.org/doc/html/draft-scudder-bgp-entropy-label-00>>.

[I-D.uttaro-idr-bgp-oad]

Uttaro, J., Retana, A., Mohapatra, P., Patel, K., and B. Wen, "One Administrative Domain using BGP", Work in Progress, Internet-Draft, draft-uttaro-idr-bgp-oad-07, 14 October 2025, <<https://datatracker.ietf.org/doc/html/draft-uttaro-idr-bgp-oad-07>>.

[RFC4786] Abley, J. and K. Lindqvist, "Operation of Anycast Services", BCP 126, RFC 4786, DOI 10.17487/RFC4786, December 2006, <<https://www.rfc-editor.org/rfc/rfc4786>>.

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, DOI 10.17487/RFC5065, August 2007, <<https://www.rfc-editor.org/rfc/rfc5065>>.

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

[RFC7705] George, W. and S. Amante, "Autonomous System Migration Mechanisms and Their Effects on the BGP AS_PATH Attribute", RFC 7705, DOI 10.17487/RFC7705, November 2015, <<https://www.rfc-editor.org/rfc/rfc7705>>.

[RFC8955] Loibl, C., Hares, S., Raszuk, R., McPherson, D., and M. Bacher, "Dissemination of Flow Specification Rules", RFC 8955, DOI 10.17487/RFC8955, December 2020, <<https://www.rfc-editor.org/rfc/rfc8955>>.

Appendix A. A Case Where a Link-Local Next Hop Could Lead to a False Positive

Consider a simple BGP peering topology, with four routers, in three Autonomous Systems:

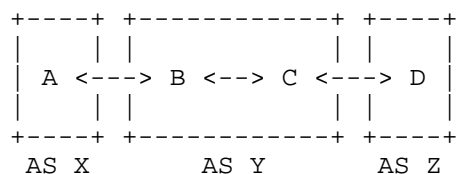


Figure 4: A Trivial Peering Topology

Suppose A and D support NHC. B and C do not support NHC. In this case, when A originates a route with an attached NHC, if B propagates it to C, and C updates the NEXT_HOP when propagating it to D, D will follow the procedures of Section 2.3 and will discard the NHC without further processing.

However, now suppose that on the peerings between A and B, and between C and D, only link-local addresses are used. Further, suppose that A uses link-local address L as its local address on its peering with B, and C also uses the same address, L, as its local address on its peering with D. In the situation described in the previous paragraph, D would have no way of detecting that C had violated the correctness assumptions of this specification, due to the collision between its address and A's.

It can be seen that since the scope of a link-local address is, of course, only the local link, the problem to be solved is restricted to knowing whether an immediate peer whose link-local address appears in the NHC is truly the originator of that NHC, or if it might be an NHC-incapable speaker that has propagated an NHC that originated elsewhere, with a colliding address.

It can further be seen that if the procedures of Section 3 are followed, this issue is resolved since A will attach a BGPID TLV containing its own BGP Identifier and its AS Number, X. Even if C's BGP Identifier is the same as A's, its AS Number is different, and thus D will discard the NHC without further processing.

Acknowledgements

The authors of this specification thank Randy Bush, Mach Chen, Wes Hardaker, Jeff Haas, Susan Hares, Ketan Talaulikar, and Gyan Mishra for their review and comments.

This specification derives from two earlier documents, [I-D.ietf-idr-next-hop-capability] and [I-D.scudder-bgp-entropy-label].

[I-D.ietf-idr-next-hop-capability] included the following acknowledgements:

The Entropy Label Next-Hop Capability defined in this document is based on the ELC BGP attribute defined in section 5.2 of [RFC6790].

The authors wish to thank John Scudder for the discussions on this topic and Eric Rosen for his in-depth review of this document.

The authors wish to thank Jie Dong and Robert Raszuk for their review and comments.

[I-D.scudder-bgp-entropy-label] included the following acknowledgements:

Thanks to Swadesh Agrawal, Alia Atlas, Bruno Decraene, Martin Djernaes, John Drake, Adrian Farrell, Keyur Patel, Toby Rees, and Ravi Singh, for their discussion of this issue.

Contributors

Wim Henderickx
Nokia
Email: wim.henderickx@nokia.com

James Uttaro
Independent Contributor
Email: juttaro@ieee.org

Authors' Addresses

Bruno Decraene (editor)
Orange
Email: bruno.decraene@orange.com

Kireeti Kompella
HPE
Email: kireeti@juniper.net

Serge Krier
Cisco Systems
Email: sekrier@cisco.com

Satya Mohanty
Zscaler
Email: smohanty@zscaler.com

John G. Scudder (editor)
HPE
Email: jgs@bgp.nu

Kevin Wang
HPE
Email: kfwang@juniper.net

Bin Wen
Comcast
Email: Bin_Wen@comcast.com