

Internet Engineering Task Force (IETF)
Internet Draft
Update: 1997, 4271, 4360, 5701 (if approved)
Intended Status: Standards Track
Expires: August 6, 2014

J. Scudder
Juniper Networks
E. Chen
Cisco Systems
P. Mohapatra
Cumulus Networks
K. Patel
Cisco Systems
February 5, 2014

Revised Error Handling for BGP UPDATE Messages
draft-ietf-idr-error-handling-05.txt

Abstract

According to the base BGP specification, a BGP speaker that receives an UPDATE message containing a malformed attribute is required to reset the session over which the offending attribute was received. This behavior is undesirable as a session reset would impact not only routes with the offending attribute, but also other valid routes exchanged over the session. This document partially revises the error handling for UPDATE messages, and provides guidelines for the authors of documents defining new attributes. Finally, it revises the error handling procedures for a number of existing attributes.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/1id-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on August 6, 2014.

Copyright Notice

Copyright (c) 2014 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 (<http://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.

1. Introduction

According to the base BGP specification [[RFC4271](#)], a BGP speaker that receives an UPDATE message containing a malformed attribute is required to reset the session over which the offending attribute was received. This behavior is undesirable as a session reset would impact not only routes with the offending attribute, but also other valid routes exchanged over the session. In the case of optional transitive attributes, the behavior is especially troublesome and may present a potential security vulnerability. The reason is that such attributes may have been propagated without being checked by intermediate routers that do not recognize the attributes -- in effect the attribute may have been tunneled, and when they do reach a router that recognizes and checks them, the session that is reset may not be associated with the router that is at fault.

The goal for revising the error handling for UPDATE messages is to minimize the impact on routing by a malformed UPDATE message, while maintaining protocol correctness to the extent possible. This can be achieved largely by maintaining the established session and keeping the valid routes exchanged, but removing the routes carried in the malformed UPDATE from the routing system.

This document partially revises the error handling for UPDATE messages, and provides guidelines for the authors of documents defining new attributes. Finally, it revises the error handling procedures for a number of existing attributes. Specifically, the error handling procedures of [[RFC4271](#)], [[RFC1997](#)], [[RFC4360](#)] and [[RFC5701](#)] are revised.

1.1. Specification of Requirements

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](#)].

2. Revision to Base Specification

The first paragraph of [Section 6.3 of \[RFC4271\]](#) is revised as follows:

Old Text:

All errors detected while processing the UPDATE message MUST be indicated by sending the NOTIFICATION message with the Error Code UPDATE Message Error. The error subcode elaborates on the specific nature of the error.

New text:

An error detected while processing the UPDATE message for which a session reset is specified MUST be indicated by sending the NOTIFICATION message with the Error Code UPDATE Message Error. The error subcode elaborates on the specific nature of the error.

The error handling of the following case described in [Section 6.3 of \[RFC4271\]](#) remains unchanged:

If the Withdrawn Routes Length or Total Attribute Length is too large (i.e., if Withdrawn Routes Length + Total Attribute Length + 23 exceeds the message Length), then the Error Subcode MUST be set to Malformed Attribute List.

The error handling of the following case described in [Section 6.3 of \[RFC4271\]](#) is revised

If any recognized attribute has Attribute Flags that conflict with the Attribute Type Code, then the Error Subcode MUST be set to Attribute Flags Error. The Data field MUST contain the erroneous attribute (type, length, and value).

as follows:

If the "Optional bit" or the "Transitive bit" in the Attribute Flags for an attribute conflicts with the Attribute Type Code, then the error SHOULD be logged, and the conflicting bit in the Attribute Flags MUST be reset to the correct value. The UPDATE

message MUST continue to be processed.

The error handling of all other cases described in [Section 6.3 of \[RFC4271\]](#) that specify a session reset is revised as follows.

When a path attribute (other than the MP_REACH_NLRI attribute [\[RFC4760\]](#) or the MP_UNREACH_NLRI attribute [\[RFC4760\]](#)) in an UPDATE message is determined to be malformed, the UPDATE message containing that attribute MUST be treated as though all contained routes had been withdrawn just as if they had been listed in the WITHDRAWN ROUTES field (or in the MP_UNREACH_NLRI attribute if appropriate) of the UPDATE message, thus causing them to be removed from the Adj-RIB-In according to the procedures of [\[RFC4271\]](#). In the case of an attribute which has no effect on route selection or installation, the malformed attribute MAY instead be discarded and the UPDATE message continue to be processed. For the sake of brevity, the former approach is termed "treat-as-withdraw", and the latter as "attribute discard".

If any of the well-known mandatory attributes are not present in an UPDATE message, then the approach of "treat-as-withdraw" MUST be used for the error handling.

The approach of "treat-as-withdraw" MUST be used for the error handling of the cases described in [Section 6.3 of \[RFC4271\]](#) that specify a session reset and involve any of the following attributes: ORIGIN, AS_PATH, NEXT_HOP, MULTI_EXIT_DISC, and LOCAL_PREF.

The approach of "attribute discard" MUST be used for the error handling of the cases described in [Section 6.3 of \[RFC4271\]](#) that specify a session reset and involve any of the following attributes: ATOMIC_AGGREGATE and AGGREGATOR.

If the MP_REACH_NLRI attribute or the MP_UNREACH_NLRI attribute appears more than once in the UPDATE message, then a NOTIFICATION message MUST be sent with the Error Subcode "Malformed Attribute List". If any other attribute appears more than once in an UPDATE message, then all the occurrences of the attribute other than the first one SHALL be discarded and the UPDATE message continue to be processed.

When multiple malformed attributes exist in an UPDATE message, if the same approach (either "session reset", or "treat-as-withdraw" or "attribute discard") is specified for the handling of these malformed attributes, then the specified approach MUST be used. Otherwise the approach with the strongest action MUST be used following the order of "session reset", "treat-as-withdraw" and "attribute discard" from the strongest to the weakest.

A document which specifies a new attribute MUST provide specifics regarding what constitutes an error for that attribute and how that error is to be handled.

Finally, we observe that in order to use the approach of "treat-as-withdraw", the entire NLRI field and/or the MP_REACH_NLRI and MP_UNREACH_NLRI attributes need to be successfully parsed. If this is not possible, the procedures of [\[RFC4271\]](#) continue to apply. Alternatively the error handling procedures specified in [\[RFC4760\]](#) for disabling a particular AFI/SAFI MAY be followed.

[3. Parsing of NLRI Fields](#)

To facilitate the determination of the NLRI field in an UPDATE with a malformed attribute, the MP_REACH_NLRI or MP_UNREACH_NLRI attribute (if present) SHALL be encoded as the very first path attribute in an UPDATE. An implementation, however, MUST still be prepared to receive these fields in any position.

If the encoding of [\[RFC4271\]](#) is used, the NLRI field for the IPv4 unicast address family is carried immediately following all the attributes in an UPDATE. When such an UPDATE is received, we observe that the NLRI field can be determined using the "Message Length", "Withdrawn Route Length" and "Total Attribute Length" (when they are consistent) carried in the message instead of relying on the length of individual attributes in the message.

[4. Operational Considerations](#)

Although the "treat-as-withdraw" error-handling behavior defined in [Section 2](#) makes every effort to preserve BGP's correctness, we note that if an UPDATE received on an IBGP session is subjected to this treatment, inconsistent routing within the affected Autonomous System may result. The consequences of inconsistent routing can include long-lived forwarding loops and black holes. While lamentable, this issue is expected to be rare in practice, and more importantly is seen as less problematic than the session-reset behavior it replaces.

When a malformed attribute is indeed detected over an IBGP session, we RECOMMEND that routes with the malformed attribute be identified and traced back to the ingress router in the network where the routes were sourced or received externally, and then a filter be applied on the ingress router to prevent the routes from being sourced or received. This will help maintain routing consistency in the network.

Even if inconsistent routing does not arise, the "treat-as-withdraw" behavior can cause either complete unreachability or sub-optimal routing for the destinations whose routes are carried in the affected UPDATE message.

Note that "treat-as-withdraw" is different from discarding an UPDATE message. The latter violates the basic BGP principle of incremental update, and could cause invalid routes to be kept. (See also [Appendix A](#).)

For any malformed attribute which is handled by the "attribute discard" instead of the "treat-as-withdraw" approach, it is critical to consider the potential impact of doing so. In particular, if the attribute in question has or may have an effect on route selection or installation, the presumption is that discarding it is unsafe, unless careful analysis proves otherwise. The analysis should take into account the tradeoff between preserving connectivity and potential side effects.

Because of these potential issues, a BGP speaker MUST provide debugging facilities to permit issues caused by a malformed attribute to be diagnosed. At a minimum, such facilities MUST include logging an error listing the NLRI involved, and containing the entire malformed UPDATE message when such an attribute is detected. The malformed UPDATE message SHOULD be analyzed, and the root cause SHOULD be investigated.

[5. Error Handling Procedures for Existing Attributes](#)

[5.1. ORIGIN](#)

The attribute is considered malformed if its length is not 1, or it has an undefined value [[RFC4271](#)].

An UPDATE message with a malformed ORIGIN attribute SHALL be handled using the approach of "treat-as-withdraw".

[5.2. AS_PATH](#)

The error conditions for the attribute have been defined in [[RFC4271](#)].

An UPDATE message with a malformed AS_PATH attribute SHALL be handled using the approach of "treat-as-withdraw".

[5.3.](#) NEXT_HOP

The error conditions for the NEXT_HOP attribute have been defined in [\[RFC4271\]](#).

An UPDATE message with a malformed NEXT_HOP attribute SHALL be handled using the approach of "treat-as-withdraw".

[5.4.](#) MULTI_EXIT_DESC

The attribute is considered malformed if its length is not 4 [\[RFC4271\]](#).

An UPDATE message with a malformed MULTI_EXIT_DESC attribute SHALL be handled using the approach of "treat-as-withdraw".

[5.5.](#) LOCAL_PREF

The attribute is considered malformed if its length is not 4 [\[RFC4271\]](#).

An UPDATE message with a malformed LOCAL_PREF attribute SHALL be handled as follows:

- o using the approach of "attribute discard" if the UPDATE message is received from an external neighbor, or

- o using the approach of "treat-as-withdraw" if the UPDATE message is received from an internal neighbor.

In addition, if the attribute is present in an UPDATE message from an external neighbor, the approach of "attribute discard" SHALL be used to handle the unexpected attribute in the message.

[5.6.](#) ATOMIC_AGGREGATE

The attribute SHALL be considered malformed if its length is not 0 [\[RFC4271\]](#).

An UPDATE message with a malformed ATOMIC_AGGREGATE attribute SHALL be handled using the approach of "attribute discard".

5.7. AGGREGATOR

The error conditions specified in [[RFC4271](#)] for the attribute are revised as follows:

The AGGREGATOR attribute SHALL be considered malformed if any of the following applies:

- o Its length is not 6 (when the "4-octet AS number capability" is not advertised to, or not received from the peer [[RFC4893](#)]).
- o Its length is not 8 (when the "4-octet AS number capability" is both advertised to, and received from the peer).

An UPDATE message with a malformed AGGREGATOR attribute SHALL be handled using the approach of "attribute discard".

5.8. Community

The error handling of [[RFC1997](#)] is revised as follows:

The Community attribute SHALL be considered malformed if its length is nonzero and is not a multiple of 4.

An UPDATE message with a malformed Community attribute SHALL be handled using the approach of "treat-as-withdraw".

5.9. Extended Community

The error handling of [[RFC4360](#)] is revised as follows:

The Extended Community attribute SHALL be considered malformed if its length is nonzero and is not a multiple of 8.

An UPDATE message with a malformed Extended Community attribute SHALL be handled using the approach of "treat-as-withdraw".

Note that a BGP speaker MUST NOT treat an unrecognized Extended Community Type or Sub-Type as an error.

5.10. IPv6 Address Specific BGP Extended Community Attribute

The error handling of [[RFC5701](#)] is revised as follows:

The IPv6 Address Specific Extended Community attribute SHALL be considered malformed if its length is nonzero and is not a multiple of 20.

An UPDATE message with a malformed IPv6 Address Specific Extended Community attribute SHALL be handled using the approach of "treat-as-withdraw".

Note that a BGP speaker MUST NOT treat an unrecognized IPv6 Address Specific Extended Community Type or Sub-Type as an error.

6. IANA Considerations

This document makes no request of IANA.

7. Security Considerations

This specification addresses the vulnerability of a BGP speaker to a potential attack whereby a distant attacker can generate a malformed optional transitive attribute that is not recognized by intervening routers (which thus propagate the attribute unchecked) but that causes session resets when it reaches routers that do recognize the given attribute type.

In other respects, this specification does not change BGP's security characteristics.

8. Acknowledgments

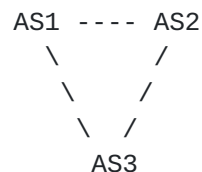
The authors wish to thank Juan Alcaide, Ron Bonica, Mach Chen, Andy Davidson, Bruno Decraene, Dong Jie, Rex Fernando, Joel Halpern, Akira Kato, Miya Kohno, Tony Li, Alton Lo, Shin Miyakawa, Tamas Mondal, Jonathan Oddy, Robert Raszuk, Yakov Rekhter, Rob Shakir, Naiming Shen, Shyam Sethuram, Ananth Suryanarayana, Kaliraj Vairavakkalai and Lili Wang for their observations and discussion of this topic, and review of this document.

9. Normative References

- [RFC1997] Chandrasekeran, R., Traina, P., and T. Li, "BGP Communities Attribute", [RFC 1997](#), August 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), January 2006.
- [RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", [RFC 4360](#), February 2006.
- [RFC4893] Vohra, Q. and E. Chen, "BGP Support for Four-octet AS Number Space", [RFC 4893](#), May 2007.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", [BCP 26](#), [RFC 5226](#), May 2008.
- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", [RFC 4760](#), January 2007.
- [RFC5701] Rekhter, Y., "IPv6 Address Specific BGP Extended Community Attribute", [RFC 5701](#), November 2009.

[Appendix A](#). Why not discard UPDATE messages?

A commonly asked question is "why not simply discard the UPDATE message instead of treating it like a withdraw? Isn't that safer and easier?" The answer is that it might be easier, but it would compromise BGP's correctness so is unsafe. Consider the following example of what might happen if UPDATE messages carrying bad attributes were simply discarded:



- o AS1 prefers to reach AS3 directly, and advertises its route to AS2.

- o AS2 prefers to reach AS3 directly, and advertises its route to AS1.
- o Connections AS3-AS1 and AS3-AS2 fail simultaneously.
- o AS1 switches to prefer AS2's route, and sends an update message which includes a withdraw of its previous announcement. The withdraw is bundled with some advertisements. It includes a bad attribute. As a result, AS2 ignores the message.
- o AS2 switches to prefer AS1's route, and sends an update message which includes a withdraw of its previous announcement. The withdraw is bundled with some advertisements. It includes a bad attribute. As a result, AS1 ignores the message.

The end result is that AS1 forwards traffic for AS3 towards AS2, and AS2 forwards traffic for AS3 towards AS1. This is a permanent (until corrected) forwarding loop.

Although the example above discusses route withdraws, we observe that in BGP the announcement of a route also withdraws the route previously advertised. The implicit withdraw can be converted into a real withdraw in a number of ways; for example, the previously-announced route might have been accepted by policy, but the new announcement might be rejected by policy. For this reason, the same concerns apply even if explicit withdraws are removed from consideration.

10. Authors' Addresses

John G. Scudder
Juniper Networks

Email: jgs@juniper.net

Enke Chen
Cisco Systems, Inc.

EMail: enkechen@cisco.com

Pradosh Mohapatra
Cumulus Networks, Inc.

EMail: pmohapat@cumulusnetworks.com

Keyur Patel
Cisco Systems, Inc.

E-Mail: keyupate@cisco.com