Internet Engineering Task Force (IETF)
Internet Draft
Update: 1997, 4271, 4360 (if approved)
Intended Status: Standards Track
Expires: April 26, 2012

J. Scudder
Juniper Networks
E. Chen
P. Mohapatra
K. Patel
Cisco Systems

October 25, 2011

Revised Error Handling for BGP UPDATE Messages
draft-ietf-idr-optional-transitive-04.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 optional attributes. Finally, it revises the error handling procedures for several existing attributes.

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 optional attributes. Finally, it revises the error handling procedures for several existing attributes. Specifically, the error handling procedures of [RFC4271], [RFC1997], and [RFC4360] 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 any attribute has Attribute Flags that conflict with the Attribute Type Code, then the error SHOULD be logged, and 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 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 [RFC4760bis] 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".

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.

When multiple malformed attributes exist in an UPDATE message, if the same approach (either "treat-as-withdraw" or "attribute discard") is specified for the handling of these malformed attributes, then the specified approach MUST be used. Otherwise "treat-as-withdraw" MUST be used.

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 MP_REACH and MP_UNREACH [RFC4760bis] 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 [RFC4760bis] 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 or MP_UNREACH attribute (if present) SHOULD be encoded as the very first path attribute in an UPDATE as recommended by [RFC4760bis]. 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 Optional Attributes

5.1. AGGREGATOR

The error handling of [RFC4271] is revised as follows:

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

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

5.2. Community

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

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

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

5.3. Extended Community

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

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

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 Ron Bonica, Mach Chen, Andy Davidson, 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, and Kaliraj Vairavakkalai 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.

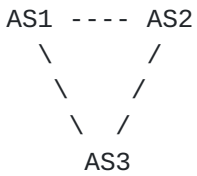
[RFC4893] Vohra, Q. and E. Chen, "BGP Support for Four-octet AS Number Space", RFC 4893, May 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC4760bis] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", draft-ietf-idr-rfc4760bis-03.txt, work in progress, August 2011.

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.

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

- Connections AS3-AS1 and AS3-AS2 fail simultaneously.

- 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.

- 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
Cisco Systems, Inc.

EMail: pmohapat@cisco.com

Keyur Patel
Cisco Systems, Inc.

EMail: keyupate@cisco.com