            Revised Error Handling for BGP UPDATE Messages
                   draft-ietf-idr-error-handling-08

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.

   This document updates error handling for RFCs 1997, 4271, 4360, 4456,
   4760 and 5701.

Status of This Memo

Copyright Notice

Table of Contents

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. To make matters
   worse, in such cases although the problematic attributes may have
   originated with a single update transmitted by a single BGP speaker,
   by the time they encounter a router that checks them they may have
   been replicated many times, and thus may cause the reset of many
   peering sessions. Thus the damage inflicted may be multiplied
   manyfold.

   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 [RFC1997], [RFC4271], [RFC4360],
   [RFC4456], [RFC4760] and [RFC5701] are revised.

1.1. Requirements Language

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. Error-Handling Approaches

In this document we refer to three different approaches to handling
errors found in BGP path attributes. They are as follows (listed in
order, from the one with the "strongest" action to the one with the
"weakest" action):

o Session reset: This is the approach used throughout the base BGP
   specification [RFC4271], where a NOTIFICATION is sent and the
   session terminated.

o Treat-as-withdraw: In this approach, the UPDATE message containing
   the path attribute in question 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].

o Attribute discard: In this approach the malformed attribute MUST
   be discarded and the UPDATE message continues to be processed.
   This approach MUST NOT be used except in the case of an attribute
   that has no effect on route selection or installation.

3. Revision to BGP UPDATE Message Error Handling

This specification amends [RFC4271] Section 6.3 in a number of ways.
See also Section 6 for treatment of specific path attributes.

a. The first paragraph 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.

b. Error handling for the following case 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.

c. Attribute Flag error handling is revised as follows:

Old Text:

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).

New Text:

If any recognized attribute has Attribute Flags that
conflict with the Attribute Type Code, then the attribute
MUST be treated as malformed and the treat-as-withdraw
approach used, unless the specification for the attribute
mandates different handling for incorrect Attribute Flags.

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

e. "Treat-as-withdraw" MUST be used for the cases that specify a
session reset and involve any of the attributes ORIGIN, AS_PATH,
NEXT_HOP, MULTI_EXIT_DISC, or LOCAL_PREF.

f. "Attribute discard" MUST be used for any of the cases that
specify a session reset and involve ATOMIC_AGGREGATE or
AGGREGATOR.

g. If the MP_REACH_NLRI attribute or the MP_UNREACH_NLRI [RFC4760]
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.

h. When multiple attribute errors exist in an UPDATE message, if the
    same approach (either "session reset", "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.

i. The Withdrawn Routes field MUST be checked for syntactic
    correctness in the same manner as the NLRI field. This is
    discussed further below, and in Section 4.2.

j. 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, meaning that the "session reset" approach SHOULD be
    followed. Alternatively the error handling procedures specified
    in [RFC4760] for disabling a particular AFI/SAFI MAY be followed.

    One notable case where it would be not possible to successfully
    parse the NLRI is if the NLRI field is found to be "syntactically
    incorrect" (see Section 4.2). It can be seen that therefore,
    this part of [RFC4271] Section 6.3 necessarily continues to
    apply:

       The NLRI field in the UPDATE message is checked for syntactic
       validity. If the field is syntactically incorrect, then the
       Error Subcode MUST be set to Invalid Network Field.

4. Parsing of NLRI Fields

To facilitate the determination of the NLRI field in an UPDATE with a
malformed attribute, the following restrictions on encoding NLRI MUST
be followed:

o The MP_REACH_NLRI or MP_UNREACH_NLRI attribute (if present) SHALL
   be encoded as the very first path attribute in an UPDATE.

o The MP_REACH_NLRI or MP_UNREACH_NLRI SHALL NOT be combined in the
   same UPDATE message.

o The MP_REACH_NLRI and MP_UNREACH_NLRI attributes MUST NOT be used
   in an UPDATE that also contains a non-empty Withdrawn Routes or
   Network Layer Reachability Information field.

In all these cases, however, an implementation MUST still be prepared
to receive these fields in any position or combination.

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.

4.1. Attribute Length Fields

There are two error cases in which the Total Attribute Length value
can be in conflict with the enclosed path attributes, which
themselves carry length values. In the "overrun" case, as the
enclosed path attributes are parsed, the length of the last
encountered path attribute would cause the Total Attribute Length to
be exceeded. In the "underrun" case, as the enclosed path attributes
are parsed, after the last successfully-parsed attribute, fewer than
three octets remain, or fewer than four octets, if the Attribute
Flags field has the Extended Length bit set -- that is, there remains
unconsumed data in the path attributes but yet insufficient data to
encode a single minimum-sized path attribute. In either of these
cases an error condition exists and the treat-as-withdraw approach
MUST be used (unless some other, more severe error is encountered
dictating a stronger approach), and the Total Attribute Length MUST
be relied upon to enable the beginning of the NLRI field to be
located.

For all path attributes other than those specified as having an
attribute length that may be zero it SHALL be considered a syntax
error for the attribute to have a length of zero. (Of the path
attributes considered in this specification, only AS_PATH and
ATOMIC_AGGREGATE may validly have an attribute length of zero.)

4.2. Syntactic Correctness of NLRI Fields

The NLRI field or Withdrawn Routes field SHALL be considered
"syntactically incorrect" if either of the following are true:

o The length of any of the included NLRI is greater than 32,

o When parsing NLRI contained in the field, the length of the last
   NLRI found exceeds the amount of unconsumed data remaining in the
   field.

Similarly, the MP_REACH or MP_UNREACH attribute of an update SHALL be
considered to be incorrect if any of the following are true:

   o The length of any of the included NLRI is inconsistent with the
      given AFI/SAFI (for example, if an IPv4 NLRI has a length greater
      than 32 or an IPv6 NLRI has a length greater than 128),

   o When parsing NLRI contained in the attribute, the length of the
      last NLRI found exceeds the amount of unconsumed data remaining in
      the attribute.

   o The attribute flags of the attribute are inconsistent with those
      specified in [RFC4760].

   o The length of the MP_UNREACH attribute is less than 3, or the
      length of the MP_REACH attribute is less than 5.

4.3. Typed NLRI

   Certain address families, for example MVPN [RFC7117] and EVPN
   [I-D.ietf-l2vpn-evpn] have NLRI that are typed. Since supported type
   values within the address family are not expressed in the MP-BGP
   capability [RFC4760], it is possible for a BGP speaker to advertise
   support for the given address family and sub-address family while
   still not supporting a particular type of NLRI within that AFI/SAFI.

   A BGP speaker advertising support for such a typed address family
   MUST handle routes with unrecognized NLRI types within that address
   family by discarding them, unless the relevant specification for that
   address family specifies otherwise.

5. 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.

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.

6. Error Handling Procedures for Existing Attributes

In the following subsections, we elaborate on the conditions for
error-checking various path attributes, and specify what approach(es)
should be used to handle malformations. It is possible that
implementations may apply other error checks not contemplated here.
If so, the error handling approach given here should generally be
applied.

6.1. 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".

6.2. AS_PATH

An AS_PATH is considered malformed if an unrecognized segment type is
encountered, or if it contains a malformed segment. A segment is
considered malformed if any of the following obtains:

o There is an overrun, where the path segment length field of the
   last segment encountered would cause the Attribute Length to be
   exceeded.

o There is an underrun, where after the last successfully-parsed
   segment, there is only a single octet remaining (that is, there is
   not enough unconsumed data to provide even an empty segment
   header).

o It has a path segment length field of zero.

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

[RFC4271] also says that an implementation optionally "MAY check
whether the leftmost ... AS in the AS_PATH attribute is equal to the
autonomous system number of the peer that sent the message". A BGP
implementation SHOULD also handle routes that violate this check
using "treat-as-withdraw", but MAY follow the session reset behavior
if configured to do so.

6.3. NEXT_HOP

The attribute is considered malformed if it is syntactically
incorrect according to [RFC4271].

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

6.4. MULTI_EXIT_DISC

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".

6.5. LOCAL_PREF

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

o If the LOCAL_PREF attribute is received from an external neighbor,
   it SHALL be discarded using the approach of "attribute discard",
   or

o if received from an internal neighbor, it SHALL be considered
   malformed if its length is not equal to 4. If malformed, the
   UPDATE SHALL be handled using the approach of "treat-as-withdraw".

6.6. 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".

6.7. 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 [RFC6793]).

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".

6.8. 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".

6.9. 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.

6.10. 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 not a nonzero 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.

6.11. ORIGINATOR_ID

The error handling of [RFC4456] is revised as follows.

o If the ORIGINATOR_ID attribute is received from an external
   neighbor, it SHALL be discarded using the approach of "attribute
   discard", or

o if received from an internal neighbor, it SHALL be considered
   malformed if its length is not equal to 4. If malformed, the
   UPDATE SHALL be handled using the approach of "treat-as-withdraw".

6.12. CLUSTER_LIST

The error handling of [RFC4456] is revised as follows.

o If the CLUSTER_LIST attribute is received from an external
   neighbor, it SHALL be discarded using the approach of "attribute
   discard", or

o if received from an internal neighbor, it SHALL be considered
   malformed if its length is not a nonzero multiple of 4. If
   malformed, the UPDATE SHALL be handled using the approach of
   "treat-as-withdraw".

6.13. MP_REACH_NLRI and MP_UNREACH_NLRI

The handling of these attributes is discussed in Section 3 and
Section 4.

7. Guidance for Authors of BGP Specifications

A document that specifies a new BGP attribute MUST provide specifics
regarding what constitutes an error for that attribute and how that
error is to be handled. Allowable error-handling approaches are
detailed in Section 2. The treat-as-withdraw approach is generally
preferred. The document SHOULD also provide consideration of what
debugging facilities may be required to permit issues caused by a
malformed attribute to be diagnosed.

For any malformed attribute that 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.

8. IANA Considerations

This document makes no request of IANA.

9. 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.

10. Acknowledgements

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

11. References

11.1. 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.

   [RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route
              Reflection: An Alternative to Full Mesh Internal BGP
              (IBGP)", RFC 4456, April 2006.

   [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.

   [RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet
              Autonomous System (AS) Number Space", RFC 6793, December
              2012.

11.2. Informative References

   [I-D.ietf-l2vpn-evpn]
              Sajassi, A., Aggarwal, R., Bitar, N., Isaac, A., and J.
              Uttaro, "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-
              evpn-07 (work in progress), May 2014.

   [RFC7117] Aggarwal, R., Kamite, Y., Fang, L., Rekhter, Y., and C.
              Kodeboniya, "Multicast in Virtual Private LAN Service
              (VPLS)", RFC 7117, February 2014.

Authors' Addresses

   Enke Chen (editor)
   Cisco Systems, Inc.

   Email: enkechen@cisco.com

   John G. Scudder (editor)
   Juniper Networks

   Email: jgs@juniper.net

   Pradosh Mohapatra
   Sproute Networks

   Email: mpradosh@yahoo.com

   Keyur Patel
   Cisco Systems, Inc.

   Email: keyupate@cisco.com