

Network Working Group  
Internet Draft  
Updates: [4271](#) (if approved)  
Intended Status: Standards Track  
Expiration Date: March 9, 2011

E. Chen  
P. Mohapatra  
K. Patel  
Cisco Systems  
September 8, 2010

**Revised Error Handling for BGP Updates from External Neighbors**  
**draft-chen-ebgp-error-handling-00.txt**

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 March 9, 2011.

Copyright Notice

Copyright (c) 2010 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.

## Abstract

In this document we partially revise the error handling of an UPDATE message from an external BGP neighbor. The essence of the revision is to avoid resetting an external BGP session by using the "treat-as-withdraw" approach when the whole NLRI field of a malformed UPDATE message can be parsed.

## 1. Introduction

The base BGP specification [[RFC4271](#)] requires that a BGP session be reset when an UPDATE message containing a malformed attribute is received. This behavior is undesirable in the case of optional transitive attributes as has been discussed and revised in [OPT-TRANS].

However, there are other situations where the behavior is also undesirable, but are outside the scope of [OPT-TRANS]. For example, there have been a few occurrences in the field where the AS-PATH attribute is malformed for a small number of routes. Resetting the BGP session would impact all the other valid routes in these cases.

Our goal is to minimize the scope of the network that is affected by a malformed UPDATE message, and also to limit the impact to only the routes involved. The constrain is that the protocol correctness must not be violated.

In this document we partially revise the error handling of an UPDATE message from an external BGP neighbor. The essence of the revision is to avoid resetting an external BGP session by using the "treat-as-withdraw" approach specified in [OPT-TRANS] when the whole NLRI field of a malformed UPDATE message can be parsed.

### 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 [[RFC2119](#)].



## **2. Revision to Base Specification**

The revised error handling specified in this section is applicable only for processing an UPDATE message from an external BGP neighbor.

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

If a path attribute in an UPDATE message from an external BGP neighbor is determined to be malformed, the message containing that attribute SHOULD be treated as though all contained routes had been withdrawn ("treat-as-withdraw") when the whole NLRI field in the message can be parsed.

One exception is that the "attribute discard" approach [[OPT-TRANS](#)] SHOULD be used to handle a malformed optional transitive attribute for which the "attribute discard" approach is specified.

A BGP speaker MUST provide debugging facilities to permit issues caused by malformed UPDATE messages to be diagnosed. At a minimum, such facilities SHOULD include logging an error when such an attribute is detected. The malformed UPDATE message SHOULD be analyzed, and the root cause SHOULD be investigated.

## **3. Parsing of NLRI Fields**

As described in [[OPT-TRANS](#)], we observe that in order to use the "treat-as-withdraw" approach for a malformed UPDATE, the NLRI field and/or MP\_REACH and MP\_UNREACH [[RFC4760](#)] attributes need to be successfully parsed. If this were not possible, the UPDATE would necessarily be malformed in some other way beyond the scope of this document and therefore, the procedures of [[RFC4271](#)] would continue to apply.

To facilitate the determination of the NLRI field in an UPDATE with malformed attributes, we strongly RECOMMEND that the MP\_REACH or MP\_UNREACH attribute (if present) be encoded as the very first path attribute in an UPDATE.



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

Furthermore, it is observed that the NLRIs for the IPv4 unicast address family can also be carried in the MP\_REACH attribute of an UPDATE when the IPV4 unicast address family capability is shared (i.e., both advertised and received) over a BGP session. For the same sake of better debugging and fault handling, we also RECOMMEND that the MP\_REACH attribute be used and be placed as the very first path attribute in an UPDATE in this case.

#### **4. Discussion**

As discussed in [[OPT-TRANS](#)], the approach of "treat-as-withdraw" is not always safe to use. In the case of internal BGP sessions, the resolution of recursive nexthops can result in forwarding loops and blakholes when the BGP speakers inside a network have inconsistent routing information.

Depending on the network topology, the routing table, routes involved, and whether "tunnels" are used inside a network, the approach of "treat-as-withdraw" may work for internal BGP sessions only in some specific cases. Thus it may be deployed for internal BGP sessions only as a temporary measure to stop continuous session flaps due to malformed UPDATE messages. Such deployment must be carefully evaluated on a case-by-case basis.

#### **5. IANA Considerations**

This document makes no request of IANA.



## **6. Security Considerations**

TBD

## **7. Acknowledgments**

We would like to thank Robert Raszuk, Naiming Shen and Tony Li for their review and discussions.

## **8. References**

### **8.1. Normative References**

- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), January 2006.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

### **8.2. Informative References**

- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", [RFC 4760](#), January 2007.
- [OPT-TRANS] Scudder, J. and E. Chen, "Error Handling for Optional Transitive BGP Attributes", Work in Progress, March 2010.

## **9. Authors' Addresses**

Enke Chen  
Cisco Systems, Inc.  
170 W. Tasman Dr.  
San Jose, CA 95134

EMail: [enkechen@cisco.com](mailto:enkechen@cisco.com)

Pradosh Mohapatra  
Cisco Systems, Inc.  
170 W. Tasman Dr.  
San Jose, CA 95134





EMail: pmohapat@cisco.com

Keyur Patel  
Cisco Systems, Inc.  
170 W. Tasman Dr.  
San Jose, CA 95134

EMail: keyupate@cisco.com