Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: November 3, 2017

K. Patel
Arrcus
R. Fernando
Cisco Systems
J. Scudder
J. Haas
Juniper Networks

May 2, 2017

Notification Message support for BGP Graceful Restart
draft-ietf-idr-bgp-gr-notification-11.txt

Abstract

The current BGP Graceful Restart mechanism limits the usage of BGP Graceful Restart to BGP protocol messages other than a BGP NOTIFICATION message. This document defines an extension to the BGP Graceful Restart that permits the Graceful Restart procedures to be performed when the BGP speaker receives a BGP NOTIFICATION Message or the Hold Time expires. This document also defines a new BGP NOTIFICATION Cease Error subcode whose effect is to request a full session restart instead of a Graceful Restart.

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 http://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 November 3, 2017.

Copyright Notice

Copyright (c) 2017 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.

Table of Contents

1. Introduction 2
1.1. Requirements Language 3
2. Modifications to BGP Graceful Restart Capability 3
3. BGP Hard Reset Subcode 4
3.1. Sending a Hard Reset 4
3.2. Receiving a Hard Reset 4
4. Operation 4
4.1. Rules for the Receiving Speaker 5
5. Acknowledgements 6
6. IANA Considerations 6
7. Security Considerations 6
8. Normative References 6
Authors' Addresses 6

1. Introduction

For many classes of errors, the BGP protocol must send a NOTIFICATION message and reset the peering session to handle the error condition. The BGP Graceful Restart extension defined in [RFC4724] requires that normal BGP procedures defined in [RFC4271] be followed when a NOTIFICATION message is sent or received. This document defines an extension to BGP Graceful Restart that permits the Graceful Restart procedures to be performed when the BGP speaker receives a NOTIFICATION message or the Hold Time expires. This permits the BGP speaker to avoid flapping reachability and continue forwarding while the BGP speaker restarts the session to handle errors detected in the BGP protocol.

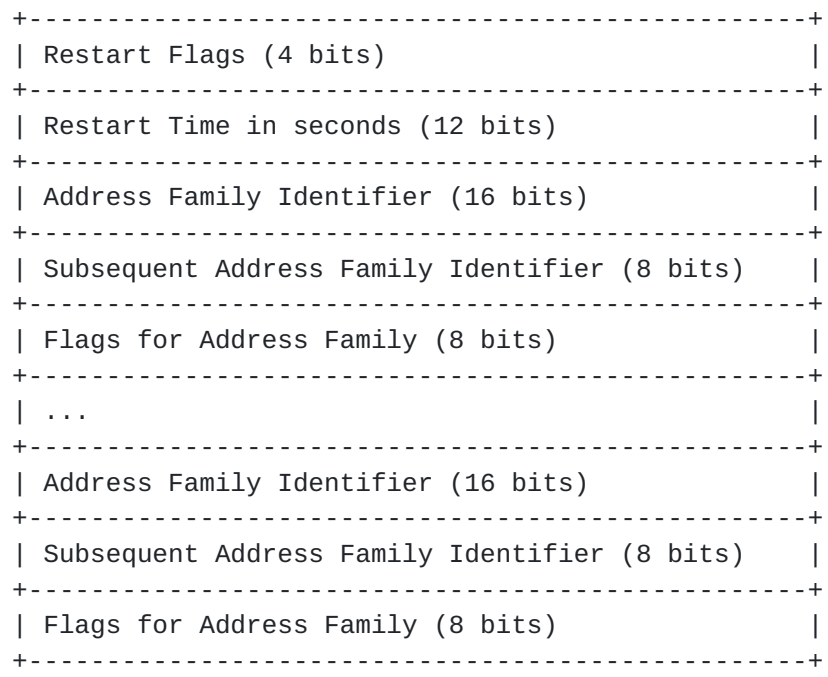
At a high level, this document can be summed up as follows. When a BGP session is reset, both speakers operate as "Receiving Speakers" according to [RFC4724], meaning they retain each other's routes. This is also true for HOLDTIME expiration. The functionality can be defeated using a "Hard Reset" subcode for the BGP NOTIFICATION Cease Error code. If a Hard Reset is used, a full session reset is performed.

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].

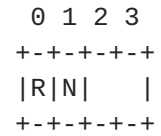
2. Modifications to BGP Graceful Restart Capability

The BGP Graceful Restart Capability is augmented to signal the Graceful Restart support for BGP NOTIFICATION messages. The Restart flags field is augmented as follows:



Restart Flags:

This field contains bit flags relating to restart.



The most significant ("Restart State", or "R") bit is defined in [RFC4724].

The second most significant bit ("N") is defined as the BGP Graceful Notification bit, which is used to indicate Graceful Restart support for BGP NOTIFICATION messages. A BGP speaker indicates support for

the procedures of this document, by advertising a Graceful Restart Capability with its Graceful NOTIFICATION bit set (value 1). This also implies support for the format for a BGP NOTIFICATION Cease message defined in [RFC4486].

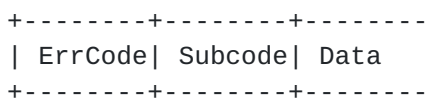
3. BGP Hard Reset Subcode

A new BGP NOTIFICATION Cease message subcode is defined known as the BGP Hard Reset Subcode. The value of this subcode is discussed in Section 6. We refer to a BGP NOTIFICATION Cease message with the Hard Reset subcode as a Hard Reset message, or just a Hard Reset.

3.1. Sending a Hard Reset

A Hard Reset message is used to indicate to a peer with which the Graceful Notification flag has been exchanged, that the session is to be fully terminated.

When sending a Hard Reset, the data portion of the NOTIFICATION is encoded as follows:



ErrCode is a BGP Error Code (as documented in the IANA BGP Error Codes registry) that indicates the reason for the hard reset. Subcode is a BGP Error Subcode (as documented in the IANA BGP Error Subcodes registry) as appropriate for the ErrCode. Similarly, Data is as appropriate for the ErrCode and Subcode.

3.2. Receiving a Hard Reset

Whenever a BGP speaker receives a Hard Reset, the speaker MUST terminate the BGP session following the standard procedures in [RFC4271].

4. Operation

A BGP speaker that is willing to receive and send BGP NOTIFICATION messages in Graceful mode MUST advertise the BGP Graceful Notification "N" bit using the Graceful Restart Capability as defined in [RFC4724].

When such a BGP speaker has received the "N" bit from its peer, and receives from that peer a BGP NOTIFICATION message other than a Hard Reset, it MUST follow the rules for the Receiving Speaker mentioned

in Section 4.1. The BGP speaker generating the BGP NOTIFICATION message MUST also follow the rules for the Receiving Speaker.

When a BGP speaker resets its session due to a HOLDTIME expiry, it should generate the relevant BGP NOTIFICATION message as mentioned in [RFC4271], but subsequently it MUST follow the rules for the Receiving Speaker mentioned in Section 4.1.

A BGP speaker SHOULD NOT send a Hard Reset to a peer from which it has not received the "N" bit. We note, however, that if it did so the effect would be as desired in any case, since according to [RFC4271] and [RFC4724] any NOTIFICATION message, whether recognized or not, results in a session reset. Thus the only negative effect to be expected from sending the Hard Reset to a peer that hasn't advertised compliance to this specification would be that the peer would be unable to properly log the associated information.

Once the session is re-established, both BGP speakers SHOULD set their "Forwarding State" bit to 1. If the "Forwarding State" bit is not set, then according to the procedures of [RFC4724] S. 4.2, the relevant routes will be flushed, defeating the goals of this specification.

4.1. Rules for the Receiving Speaker

[RFC4724] S. 4.2 defines rules for the Receiving Speaker. These are modified as follows.

As part of this extension, routes from the peer previously marked as stale MUST NOT be deleted, until and unless the optional timer mentioned in the final paragraph of [RFC4724] S. 4.2 expires, or unless a Hard Reset is performed. This supersedes the "consecutive restarts" requirement in the third paragraph of [RFC4724] S. 4.2.

In addition to the rules already specified in [RFC4724] S. 4.2 for how variations in the received Graceful Restart Capability should be interpreted (the paragraph that begins "Once the session is re-established..."), if the Graceful Notification ("N") bit is not set in the newly received Graceful Restart Capability, no new actions are triggered on the Receiving Speaker -- in particular, a clear "N" bit does not trigger deletion of stale routes.

Other than these modifications, the rules for the Receiving Speaker are as specified in [RFC4724] S. 4.2.

5. Acknowledgements

The authors would like to thank Jim Uttaro for the suggestion, and Emmanuel Baccelli, Bruno Decraene, Chris Hall, Paul Mattes and Robert Raszuk for their review and comments.

6. IANA Considerations

IANA has temporarily assigned subcode 9, named "Hard Reset", in the "BGP Cease NOTIFICATION message subcodes" registry. Upon publication of this document as an RFC, IANA is requested to make this allocation permanent.

7. Security Considerations

This extension to BGP does not change the underlying security issues inherent in the existing [RFC4724] and [RFC4271].

8. 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, <http://www.rfc-editor.org/info/rfc2119>.

[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, <http://www.rfc-editor.org/info/rfc4271>.

[RFC4486] Chen, E. and V. Gillet, "Subcodes for BGP Cease Notification Message", RFC 4486, DOI 10.17487/RFC4486, April 2006, <http://www.rfc-editor.org/info/rfc4486>.

[RFC4724] Sangli, S., Chen, E., Fernando, R., Scudder, J., and Y. Rekhter, "Graceful Restart Mechanism for BGP", RFC 4724, DOI 10.17487/RFC4724, January 2007, <http://www.rfc-editor.org/info/rfc4724>.

Authors' Addresses

Keyur Patel
Arrcus

Email: keyur@arrcus.com

Rex Fernando
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: rex@cisco.com

John Scudder
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
USA

Email: jgs@juniper.net

Jeff Haas
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
USA

Email: jhaas@juniper.net