Internet Engineering Task Force (IETF)
Request for Comments: 7911
Category: Standards Track
ISSN: 2070-1721

D. Walton
Cumulus Networks
A. Retana
E. Chen
Cisco Systems, Inc.
J. Scudder
Juniper Networks

July 2016

Advertisement of Multiple Paths in BGP

Abstract

This document defines a BGP extension that allows the advertisement of multiple paths for the same address prefix without the new paths implicitly replacing any previous ones. The essence of the extension is that each path is identified by a Path Identifier in addition to the address prefix.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7911.

Table of Contents

1. Introduction

The BGP specification [RFC4271] defines an Update-Send Process to advertise the routes chosen by the Decision Process to other BGP speakers. No provisions are made to allow the advertisement of multiple paths for the same address prefix or Network Layer Reachability Information (NLRI). In fact, a route with the same NLRI as a previously advertised route implicitly replaces the previous advertisement.

This document defines a BGP extension that allows the advertisement of multiple paths for the same address prefix without the new paths implicitly replacing any previous ones. The essence of the extension is that each path is identified by a Path Identifier in addition to the address prefix.

The availability of the additional paths can help reduce or eliminate persistent route oscillations [RFC3345]. It can also help with optimal routing and routing convergence in a network by providing potential alternate or backup paths, respectively.

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. How to Identify a Path

As defined in [RFC4271], a path refers to the information reported in the Path Attribute field of an UPDATE message. As the procedures specified in [RFC4271] allow only the advertisement of one path for a particular address prefix, a path for an address prefix from a BGP peer can be keyed on the address prefix.

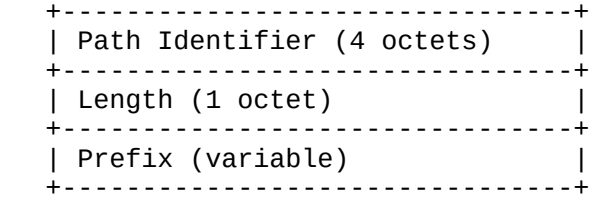
In order for a BGP speaker to advertise multiple paths for the same address prefix, a new identifier (termed "Path Identifier" hereafter) needs to be introduced so that a particular path for an address prefix can be identified by the combination of the address prefix and the Path Identifier.

The assignment of the Path Identifier for a path by a BGP speaker is purely a local matter. However, the Path Identifier MUST be assigned in such a way that the BGP speaker is able to use the (Prefix, Path Identifier) to uniquely identify a path advertised to a neighbor. A BGP speaker that re-advertises a route MUST generate its own Path Identifier to be associated with the re-advertised route. A BGP speaker that receives a route should not assume that the identifier carries any particular semantics.

3. Extended NLRI Encodings

In order to carry the Path Identifier in an UPDATE message, the NLRI encoding MUST be extended by prepending the Path Identifier field, which is of four octets.

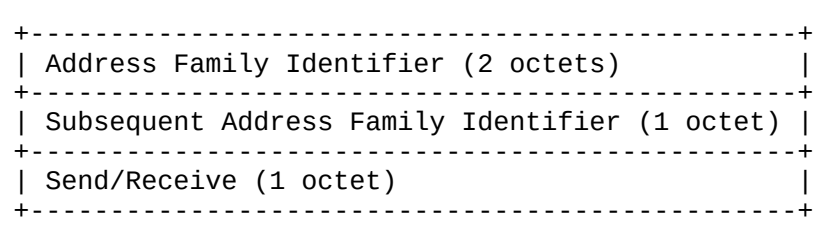
For example, the NLRI encoding specified in [RFC4271] is extended as the following:



The usage of the extended NLRI encodings is specified in Section 5.

4. ADD-PATH Capability

The ADD-PATH Capability is a BGP capability [RFC5492], with Capability Code 69. The Capability Length field of this capability is variable. The Capability Value field consists of one or more of the following tuples:



The meaning and use of the fields are as follows:

Address Family Identifier (AFI):

This field is the same as the one used in [RFC4760].

Subsequent Address Family Identifier (SAFI):

This field is the same as the one used in [RFC4760].

Send/Receive:

This field indicates whether the sender is (a) able to receive multiple paths from its peer (value 1), (b) able to send multiple paths to its peer (value 2), or (c) both (value 3) for the <AFI, SAFI>.

If any other value is received, then the capability SHOULD be treated as not understood and ignored [RFC5492].

A BGP speaker that wishes to indicate support for multiple AFI/SAFIs MUST do so by including the information in a single instance of the ADD-PATH Capability.

5. Operation

The Path Identifier specified in Section 3 can be used to advertise multiple paths for the same address prefix without subsequent advertisements replacing the previous ones. Apart from the fact that this is now possible, the route advertisement rules of [RFC4271] are not changed. In particular, a new advertisement for a given address prefix and a given Path Identifier replaces a previous advertisement

for the same address prefix and Path Identifier. If a BGP speaker receives a message to withdraw a prefix with a Path Identifier not seen before, it SHOULD silently ignore it.

For a BGP speaker to be able to send multiple paths to its peer, that BGP speaker MUST advertise the ADD-PATH Capability with the Send/Receive field set to either 2 or 3, and MUST receive from its peer the ADD-PATH Capability with the Send/Receive field set to either 1 or 3, for the corresponding <AFI, SAFI>.

A BGP speaker MUST follow the procedures defined in [RFC4271] when generating an UPDATE message for a particular <AFI, SAFI> to a peer unless the BGP speaker advertises the ADD-PATH Capability to the peer indicating its ability to send multiple paths for the <AFI, SAFI>, and also receives the ADD-PATH Capability from the peer indicating its ability to receive multiple paths for the <AFI, SAFI>, in which case the speaker MUST generate a route update for the <AFI, SAFI> based on the combination of the address prefix and the Path Identifier, and use the extended NLRI encodings specified in this document. The peer SHALL act accordingly in processing an UPDATE message related to a particular <AFI, SAFI>.

A BGP speaker SHOULD include the best route [RFC4271] when more than one path is advertised to a neighbor, unless it is a path received from that neighbor.

As the Path Identifiers are locally assigned, and may or may not be persistent across a control plane restart of a BGP speaker, an implementation SHOULD take special care so that the underlying forwarding plane of a "Receiving Speaker" as described in [RFC4724] is not affected during the graceful restart of a BGP session.

6. Deployment Considerations

The extension proposed in this document provides a mechanism for a BGP speaker to advertise multiple paths over a BGP session. Care needs to be taken in its deployment to ensure consistent routing and forwarding in a network [ADDPATH].

The only explicit indication that the encoding described in Section 3 is in use in a particular BGP session is the exchange of Capabilities described in Section 4. If the exchange is successful [RFC5492], then the BGP speakers will be able to process all BGP UPDATES properly, as described in Section 5. However, if, for example, a packet analyzer is used on the wire to examine an active BGP session, it may not be able to properly decode the BGP UPDATES because it lacks prior knowledge of the exchanged Capabilities.

When deployed as a provider edge router or a peering router that interacts with external neighbors, a BGP speaker usually advertises at most one path to the internal neighbors in a network. In the case where the speaker is configured to advertise multiple paths to the internal neighbors, and additional information is needed for the application, the speaker could use attributes such as the Edge_Discriminator attribute [FAST]. The use of that type of additional information is outside the scope of this document.

7. IANA Considerations

IANA has assigned the value 69 for the ADD-PATH Capability described in this document. This registration is in the "Capability Codes" registry.

8. Security Considerations

This document defines a BGP extension that allows the advertisement of multiple paths for the same address prefix without the new paths implicitly replacing any previous ones. As a result, multiple paths for a large number of prefixes may be received by a BGP speaker, potentially depleting memory resources or even causing network-wide instability, which can be considered a denial-of-service attack. Note that this is not a new vulnerability, but one that is present in the base BGP specification [RFC4272].

The use of the ADD-PATH Capability is intended to address specific needs related to, for example, eliminating route oscillations that were induced by the MULTI_EXIT_DISC (MED) attribute [STOP-OSC]. While describing the applications for the ADD-PATH Capability is outside the scope of this document, users are encouraged to examine their behavior and potential impact by studying the best practices described in [ADDPATH].

Security concerns in the base operation of BGP [RFC4271] also apply.

9. References

9.1. 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>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <http://www.rfc-editor.org/info/rfc4760>.

[RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <http://www.rfc-editor.org/info/rfc5492>.

9.2. Informative References

[ADDPATH] Uttaro, J., Francois, P., Patel, K., Haas, J., Simpson, A., and R. Fragassi, "Best Practices for Advertisement of Multiple Paths in IBGP", Work in Progress, draft-ietf-idr-add-paths-guidelines-08, April 2016.

[FAST] Mohapatra, P., Fernando, R., Filsfils, C., and R. Raszuk, "Fast Connectivity Restoration Using BGP Add-path", Work in Progress, draft-pmohapat-idr-fast-conn-restore-03, January 2013.

[RFC3345] McPherson, D., Gill, V., Walton, D., and A. Retana, "Border Gateway Protocol (BGP) Persistent Route Oscillation Condition", RFC 3345, DOI 10.17487/RFC3345, August 2002, <http://www.rfc-editor.org/info/rfc3345>.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <http://www.rfc-editor.org/info/rfc4272>.

[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>.

[STOP-OSC] Walton, D., Retana, A., Chen, E., and J. Scudder, "BGP Persistent Route Oscillation Solutions", Work in Progress, draft-ietf-idr-route-oscillation-stop-03, April 2016.

Acknowledgments

We would like to thank David Cook and Naiming Shen for their contributions to the design and development of the extension.

Many people have made valuable comments and suggestions, including Rex Fernando, Eugene Kim, Danny McPherson, Dave Meyer, Pradosh Mohapatra, Keyur Patel, Robert Raszuk, Eric Rosen, Srihari Sangli, Dan Tappan, Mark Turner, Jeff Haas, Jay Borkenhagen, Mach Chen, Denis Ovsienko, Carlos Pignataro, Meral Shirazipour, and Kathleen Moriarty.

Authors' Addresses

Daniel Walton
Cumulus Networks
185 E. Dana Street
Mountain View, CA 94041
United States of America

Email: dwalton@cumulusnetworks.com

Alvaro Retana
Cisco Systems, Inc.
Kit Creek Rd.
Research Triangle Park, NC 27709
United States of America

Email: aretana@cisco.com

Enke Chen
Cisco Systems, Inc.
170 W. Tasman Dr.
San Jose, CA 95134
United States of America

Email: enkechen@cisco.com

John Scudder
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
United States of America

Email: jgs@juniper.net