BGP Remote-Next-Hop
draft-vandevelde-idr-remote-next-hop-00

Abstract

The BGP Remote-Next-Hop optional transitive attribute, provides an additional level of recursive route-lookup. The BGP Remote-Next-Hop attribute can be used, both on the public Internet infrastructure and within public or private Autonomous Systems (AS). The usage of BGP Remote-Next-Hop operates as ships-in-the-night and does not require any capability exchange for any BGP Session. BGP Remote-Next-Hop is providing the distribution of one or more Location Identifiers assigned to a particular BGP prefix or an NLRI. To secure the BGP prefix or NLRI entry BGP security mechanisms, either RPKI or other BGP mechanisms can be utilized.

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 25, 2012.

Copyright Notice

Copyright (c) 2012 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
publications 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 .................................................. 3
   1.1. Requirements Language ................................. 3
   1.2. Option Format: BGP Remote-Next-Hop .................. 3
       1.2.1. BGP Option Type .................................. 3
       1.2.2. Option Format ..................................... 4
       1.2.3. Remote-Next-Hop Attribute Format ................. 4
   1.3. Usage Guidelines ....................................... 5
       1.3.1. Networks that do NOT support BGP Next-Hop-Attribute .. 5
       1.3.2. Multi-homing for IPv6 ............................. 5
       1.3.3. Mapping Technology to compliment LISP .......... 6
       1.3.4. Network Overlay Infrastructure .................. 6
       1.3.5. Reduction of the Routing Forwarding Information Base ................................................. 6
2. IANA Considerations ........................................... 6
3. Security Considerations .................................... 7
   3.1. Privacy Considerations ............................... 7
4. Acknowledgements ............................................ 7
5. Change Log .................................................. 7
6. References .................................................. 7
   6.1. Normative References .................................. 7
   6.2. Informative References ............................... 7
Authors’ Addresses .............................................. 8
1. Introduction

Each IP address (v4 or v6) represents a location and an identity. This is captured and developed by Locator Identity Split Protocol (LISP) and additionally documented by rfc6227 (Design Goals for Scalable Internet Routing) to scale the Internet.

The optional transitive BGP Remote-Next-Hop attribute provides a mapping to complement and support mapping technologies (i.e. LISP) by using BGP to distribute either a single or a set of locators attached to each entry in the BGP table. Based upon this locater information (the BGP Remote-Next-Hop) an overlay tunnel can be utilized and created. This overlay tunnel could be any type of IPv4-in-IPv4, IPv6-in-IPv4, IPv4-in-IPv6 or IPv6-in-IPv6.

If a recipient node has no awareness or does not understand the BGP Remote-Next-Hop attribute then traditional BGP routing can be used as fallback mechanism. If the recipient node does understand the BGP Remote-Next-Hop attribute, then an overlay tunnel could be created i.e a LISP or IPoGRE tunnel.

In addition, existing BGP security mechanisms can be used to verify the correctness of the BGP update including the validity of the BGP Remote-Next-Hop attribute. This will make sure that recipients of the BGP NLRI update which understand BGP Remote-Next-Hop, are not by accident tunneling to a malicious intermediate hop instead of the rightful BGP end-point.

This note is defining the attribute for usage on Inter-AS and Intra-AS environments.

Note that the Address Family (AF) of the NLRI exchanged is intended to be independent of the Address Family of the BGP Remote-Next-Hop attribute; hence the BGP Remote-Next-Hop attribute can be used within and between AF environments.

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 [RFC2119].

1.2. Option Format: BGP Remote-Next-Hop

1.2.1. BGP Option Type

Optional Transitive
1.2.2. Option Format

| Bit 0: Set to 1 (Optional Attribute); |
| Bit 1: Set to 1 (Transitional Attribute); |
| Bit 2: Partial bit of the attribute; |
Depending on the size and entries within the BGP Remote-Next-Hop the attribute may be complete (set to 0) or partial (set to 1); |
| Bit 3: Extended Length bit; |
Set to 0 if non-extended; |
Set to 1 if Extended length; |
| Bit 4 to 7 Set to 0; |
| Bit 8 to 15: Set to BGP Remote-Next-Hop attribute. Value to be defined by IANA; |

1.2.3. Remote-Next-Hop Attribute Format

The general format of this attribute describes the structure of the BGP Remote-Next-Hop attribute. Each attribute address family MUST be carried in a different Remote-Next-Hop container. (For example IPv4 and IPv6 will each need their own container)

Each attribute can carry multiple BGP Remote-Next-Hop addresses.
The coding mentioned here is Hexadecimals.

If Address Family is set to 0 then it are IPv4 addresses.
If Address Family is set to 1 then it are IPv6 addresses.
Other values are reserved for the future purposes.
The length field is encoded in number of Octets.


1.3. Usage Guidelines

This section provides a short overview of some use-case scenarios for the BGP Remote-Next-Hop attribute. The usage of the BGP Remote-Next-Hop attribute is not limited to the examples discussed in this section.

1.3.1. Networks that do NOT support BGP Next-Hop-Attribute

If a device does not support this attribute, then due to the Optional Transitive BGP character of this attribute will result that traditional routing is utilized with all the benefits and drawbacks.

1.3.2. Multi-homing for IPv6

When an IPv6 network is multi-homed towards the Internet, then it may be assigned with more than a single prefix. While the aspiration for these different assignments is to reduce the size of the Global Internet Routing table, it also results in well known resiliency issues discussed in a few IETF Working groups (if attribute idea gets accepted by WG, references will be added).

The BGP Remote-Next-Hop attribute is providing a technology to glue these assigned IPv6 prefixes together and avoid some of the well known resiliency issues.

So, how does this concept work?

Assume you have a Node (=Source-Node) on Network (=Source-Network) and you would like to send something to your peer (Destination-Node) which is a device in Network (Destination-Network). With traditional routing a destination lookup is done, and forwarded each hop between
Source-Node and Destination-node.

What the attribute BGP Remote-Next-Hop provides is that when the Source-Node sends a packet to the Destination-Node it can be specially handled by the Source-Network edge-router. This device will look at the destination address of the packet, and check the BGP Remote-Next-Hop attribute (assuming that the edge-node supports that idea). That lookup will allow a device understanding the attribute, to select or create a tunnel towards the destination.

Using this will result in having organizations built Internet overlays.

1.3.3. Mapping Technology to compliment LISP

LISP, the Locator ID Split protocol, makes usage of a mapping technology to build up the tunnels to accommodate LISP. What this BGP Remote-Next-Hop attribute is providing is that it provides a mechanism to distribute tunnel end-points using existing BGP infrastructure without the need of anybody requiring to enable it. As result it will be ship in the night from both BGP as LISP perspective.

1.3.4. Network Overlay Infrastructure

With the BGP Remote-Next-Hop feature it is possible to build and create an overlay tunneled network. By using this attribute, it is possible to have individual (Internet) Networks create sessions between them and make sure that the Internet and serviceability elements are kept correctly.

1.3.5. Reduction of the Routing Forwarding Information Base

Right now there are about 50,000 Autonomous Systems distributed, which is way less as the useful entries in the existing FIB. The usage of double recursive reduction has as result that even if the routing table stays equal size, that the FIB (Traditionally ASIC Based Forwarding Table) when everybody is using this technology is reduced. (We would go from 400K entries to 50k entries in the FIB).

2. IANA Considerations

This memo asks the IANA for a new BGP Attribute assignment.
3. Security Considerations

This technology could be used as technology as man in the middle attack, however with existing RPKI validation for BGP that risk is reduced.

Additional risks are still to be analysed by the working group and feedback received on the draft

3.1. Privacy Considerations

This proposal could introduce privacy issues, however with BGP security mechanisms in place they are removed.

4. Acknowledgements

To be completed

5. Change Log

Initial Version: 16 May 2012

6. References

6.1. Normative References


6.2. Informative References

Authors’ Addresses

Gunter Van de Velde
Cisco Systems
De Kleetlaan 6a
Diegem 1831
Belgium

Phone: +32 2704 5473
Email: gvandeve@cisco.com

Keyur Patel
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95124 95134
USA

Phone: +32 2704 5473
Email: keyupate@cisco.com