BGP Next-Hop dependant capabilities
draft-ietf-idr-next-hop-capability-00

Abstract

   RFC 5492 defines capabilities advertisement for the BGP peer. In
   addition, it is useful to be able to advertise BGP Next-Hop dependant
   capabilities, in particular for forwarding plane features. RFC 5492
   is not applicable because the BGP peer may be different from the BGP
   Next-Hop, in particular when BGP Route Reflection is used. This
   document defines a mechanism to advertise such BGP Next Hop dependant
   Capabilities.

   This document defines a new BGP non-transitive attribute to carry
   Next-Hop Capabilities. This attribute is deleted or possibly
   modified when the BGP Next Hop is changed.

   This document also defines a Next-Hop capability to advertise the
   ability to handle the MPLS Entropy Label defined in RFC 6790. It
   updates RFC 6790 with regard to this BGP signaling.

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].

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 December 18, 2017.

Copyright Notice

Table of Contents

1. Introduction

[RFC5492] defines capabilities advertisement for the BGP peer. In
addition, it is useful to be able to advertise BGP Next-Hop dependant
capabilities, in particular for forwarding plane features. RFC 5492
is not applicable because the BGP peer may be different from the BGP
Next-Hop, in particular when BGP Route Reflection is used. This

document defines a mechanism to advertise such BGP Next Hop
Capabilities.

This document defines a new BGP non-transitive attribute to carry
Next-Hop Capabilities. When the BGP Next Hop is changed, this
attribute is deleted or possibly modified to take into account the
capabilities of the new BGP Next-Hop. Hence it allows advertising
capabilities which are dependent of the BGP Next-Hop.

This attribute advertises the capabilities of the BGP Next-Hop for
the NLRI advertised in the same BGP update. A BGP Next-Hop may
advertise different capabilities for different set of NLRI.

This document also defines a first application to advertise the
capability to handle the MPLS Entropy Label defined in [RFC6790].
Note that RFC 6790 had originally defined a BGP attribute for this
but it has been latter deprecated in [RFC7447].

2. BGP Next-Hop dependant Capabilities Attribute

2.1. Encoding

The BGP Next-Hop dependant Capabilities Attribute is an optional,
non-transitive BGP Attribute, of value TBD1. The attribute consists
of a set of Next-Hop Capabilities.

The inclusion of a Next-Hop Capability "X" in a BGP UPDATE message,
indicates that the BGP Next-Hop, encoded in either the NEXT_HOP
attribute defined in [RFC4271] or the Network Address of Next Hop
field of the MP_REACH_NLRI attribute defined in [RFC4760], supports
the capability "X" for the NLRI advertised in this BGP UPDATE.

This document does not make a distinction between these two Next-Hop
fields and uses the term 'BGP Next-Hop' to refer to whichever one is
used in a given BGP UPDATE message.

A Next-Hop Capability is a triple (Capability Code, Capability
Length, Capability Value) aka a TLV:

```
              +-----------------------------+
              | Capability Code (1 octet)   |
              +-----------------------------+
              | Capability Length (1 octet) |
              +-----------------------------+
              | Capability Value (variable) |
              ~                             ~
              +-----------------------------+
```

Figure 1: BGP Next-Hop Capability

Capability Code: a one-octet unsigned binary integer which indicates the type of "Next-Hop Capability" advertised and unambiguously identifies an individual capability.

Capability Length: a one-octet unsigned binary integer which indicates the length, in octets, of the Capability Value field. A length of 0 indicates that no Capability Value Field is present.

Capability Value: a variable-length field from 0 to 255 octets. It is interpreted according to the value of the Capability Code.

BGP speakers SHOULD NOT include more than one instance of a Next-Hop capability with the same Capability Code, Capability Length, and Capability Value. Note, however, that processing of multiple instances of such capability does not require special handling, as additional instances do not change the meaning of the announced capability; thus, a BGP speaker MUST be prepared to accept such multiple instances.

BGP speakers MAY include more than one instance of a capability (as identified by the Capability Code) with non-zero Capability Length field, but with different Capability Value and either the same or different Capability Length. Processing of these capability instances is specific to the Capability Code and MUST be described in the document introducing the new capability.

2.2. Attribute Operation

The BGP Next-Hop dependant Capabilities attribute being non-transitive, as per [RFC4271], a BGP speaker which does not understand it will quietly ignore it and not pass it along to other BGP peers.

A BGP speaker that understands the BGP Next-Hop dependant Capabilities Attribute and does not change the BGP Next-Hop, SHOULD NOT change the BGP Next-Hop dependant Capabilities Attribute and SHOULD pass the attribute unchanged along to other BGP peers.

A BGP speaker that understands the BGP Next-Hop dependant
Capabilities Attribute and changes the BGP Next-Hop, MUST remove the
received BGP Next-Hop dependant Capabilities Attribute before
propagating the BGP UPDATE to other BGP peers. It MAY attach a new
BGP Next-Hop dependant Capabilities attribute describing the
capabilities of the new BGP Next-Hop for these NLRIs.

2.3. Capability Code Operation

A BGP speaker receiving a BGP Next-Hop Capability Code that it
supports behave as defined in the document defining this Capability
Code. A BGP speaker receiving a BGP Next-Hop Capability Code that it
does not support MUST ignore this BGP Next-Hop Capability Code. In
particular, this MUST NOT be handled as an error. In both cases, the
BGP speaker MUST examine the remaining BGP Next-Hop Capability
Code(s) that may be present in the BGP Next-Hop Capabilities
Attribute.

The BGP Next-Hop Capability Code MUST reflect the capability of the
router indicated in the BGP Next-Hop, for the NLRI advertised in the
BGP UPDATE. If a BGP speaker sets the BGP Next-Hop to an address of
a different router (e.g. R), it MUST NOT advertise BGP Next-Hop
Capabilities not supported by this router R for these NLRI.

The presence of a Next-Hop Capability SHOULD NOT influence route
selection or route preference of a route, unless tunneling is used to
reach the BGP Next-Hop or the selected route has been learnt from
EBGP (i.e. the Next-Hop is in a different AS). Indeed, it is in
general impossible for a node to know that all BGP routers of the
Autonomous System (AS) will understand a given Next-Hop Capability;
and having different routers, within an AS, use a different
preference for a route, may result in forwarding loops if tunnelling
is not used to reach the BGP Next-Hop.

An implementations MAY allow, by configuration, removing this
attribute or specific Next-Hop capabilities when advertising the
routes over EBGP.

2.4. Attribute Error Handling

A BGP Next-Hop dependant Capabilities Attribute is considered
malformed if the length of the Attribute is not equal to the sum of
all (BGP Next-Hop dependant Capability Length +2) of the capabilities
carried in this attribute. Note that "2" is the length of the fields
"Type" and "Length" of each BGP Next Hop dependant Capability, as the
capability length only account for the length of the Value field.

A document that specifies a new Next-Hop Capability SHOULD provide
specifics regarding what constitutes an error for that Next-Hop
Capability.

A BGP UPDATE message with a malformed BGP Next-Hop dependant
Capabilities Attribute SHALL be handled using the approach of
"attribute discard" defined in [RFC7606].

Unknown Next-Hop Capabilities Codes MUST NOT be considered as an
error. They MUST be silently ignored.

If a Next-Hop dependant Capability is malformed, this Next-Hop
Capability Type MUST be ignored. Others Next-Hop Capabilities MUST
be processed as usual.

3. Entropy Label Next-Hop dependant Capability

The Entropy Label Next-Hop Capability has type code 1 and a length of
0 or 1 octet.

The inclusion of the "Entropy Label" Next-Hop Capability indicates
that the BGP Next-Hop can be sent packets, for all routes indicated
in the NRLI, with a MPLS entropy label (ELI, EL) added immediately
after the label stack advertised with the NLRI.

On the receiving side, suppose BGP speaker S has determined that
packet P is to be forwarded according to BGP route R, where R is a
route of one of the labeled address families. And suppose that L is
the label stack embedded in the NLRI of route R. Then to forward
packet P according to route R, S either replaces P's top label with
L, or else pushes L onto the MPLS label stack. If the EL-Capability
is advertised in the BGP UPDATE advertising this route R, S knows
that it may safely place the ELI and an EL on the label stack
immediately beneath L.

A BGP speaker S that sends an UPDATE with the BGP Next-Hop "NH" MAY
include the Entropy Label Next-Hop Capability only if the NLRI are
labelled and for all the NLRI in the BGP UPDATE, either of the
following is true:

o Egress case: NH is the egress of the LSP advertised with the NLRI
   and its capable of handling the ELI during the lookup of the MPLS
   top label.

o Transit LSR case: NH is a transit LSR for the LSP advertised with
   the NLRI (i.e. NH swaps one of the label advertised in the NLRI)
   and next downstream BGP Next-Hop(s) has(have) advertised the
   Entropy Label Next-Hop Capability (or a similar capability

signalled by protocol P if the route is redistributed, by NH, from
protocol P to BGP).

3.1. Entropy Label Next-Hop Capability error handling

If the Entropy Label Next-Hop Capability is present more than once,
it MUST be considered as received once with a length of 0.

If the Entropy Label Next-Hop Capability is received with a length
other than 0 or 1, it is not considered malformed, but its semantics
are exactly the same as if it had a length of 1. In other words,
additional octets MUST be ignored. This is to allow for graceful
future extension.

4. IANA Considerations

4.1. Next-Hop Capabilities Attribute

IANA is requested to allocate a new Path Attribute, called "Next-Hop
Capabilities", type Code TBD1, from the "BGP Path Attributes"
registry.

4.2. Next-Hop Capability registry

The IANA is requested to create and maintain a registry entitled
"Next-Hop Capabilities".

The registration policies [RFC5226] for this registry are:

```
     1-63   IETF Review
    64-127  First Come First Served
   128-250  Standards Action
   251-254  Experimental Use
       255  Reserved
```

IANA is requested to make the following initial assignments:

          Registry Name: Next-Hop Capability.

Value	Meaning	Reference
0	Reserved (not to be allocated)	This document
1	Entropy Label	This document
2-250	Unassigned	
251-254	Experimental	This document
255	Reserved (for futur registry extension)	This document

5. Security Considerations

This document does not introduce new security vulnerabilities in BGP.
Specifically, an operator who is relying on the information carried
in BGP must have a transitive trust relationship back to the source
of the information. Specifying the mechanism(s) to provide such a
relationship is beyond the scope of this document. Please refer to
the Security Considerations section of [RFC4271] for security
mechanisms applicable to BGP.

6. Acknowledgement

The Entropy Label Next-Hop Capability defined in this document is
based on the ELC BGP attribute defined in section 5.2 of [RFC6790].

The authors wish to thank John Scudder for the discussions on this
topic and Eric Rosen for his in-depth review of this document.

The authors wish to thank Jie Dong for his review and comments.

7. References

7.1. 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>.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
           IANA Considerations Section in RFCs", BCP 26, RFC 5226,
           DOI 10.17487/RFC5226, May 2008,
           <http://www.rfc-editor.org/info/rfc5226>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and
           L. Yong, "The Use of Entropy Labels in MPLS Forwarding",
           RFC 6790, DOI 10.17487/RFC6790, November 2012,
           <http://www.rfc-editor.org/info/rfc6790>.

   [RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K.
              Patel, "Revised Error Handling for BGP UPDATE Messages",
              RFC 7606, DOI 10.17487/RFC7606, August 2015,
              <http://www.rfc-editor.org/info/rfc7606>.

7.2. Informative References

   [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>.

   [RFC7447] Scudder, J. and K. Kompella, "Deprecation of BGP Entropy
              Label Capability Attribute", RFC 7447,
              DOI 10.17487/RFC7447, February 2015,
              <http://www.rfc-editor.org/info/rfc7447>.

Authors' Addresses

   Bruno Decraene
   Orange

   Email: bruno.decraene@orange.com

   Kireeti Kompella
   Juniper Networks, Inc.
   1194 N. Mathilda Avenue
   Sunnyvale, CA 94089
   USA

   Email: kireeti.kompella@gmail.com

   Wim Henderickx
   Nokia
   Copernicuslaan 50
   Antwerp 2018, CA 95134
   Belgium

   Email: wim.henderickx@nokia.com