             Use of BGP for dissemination of ILA mapping information
                      draft-lapukhov-bgp-ila-afi-01

Abstract

   Identifier-Locator Addressing [I-D.herbert-nvo3-ila] relies on
   splitting the 128-bit IPv6 address into identifier and locator parts
   to implement identifier mobility, and network virtualization. This
   document proposes a method for distributing the identifier to locator
   mapping information using Multiprotocol Extensions for BGP-4
   [RFC4760].

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 September 22, 2016.

Copyright Notice

Table of Contents

1. Introduction

Under the ILA proposal, the IPv6 address is split in 64-bit
identifier (lower address bits) and locator (higher address bits)
portions. The locator part is determined dynamically from a mapping
table that maintains associations between the location-independent
identifiers and topologically significant locators. The hosts that
collectively implement and maintain such mappings are referred to as
"ILA domain" in this document. The ILA domain has a globally unique
64-bit SIR (Standard Identifier Representation) prefix assigned to it
(see [I-D.herbert-nvo3-ila] for more details on SIR prefix use).

This document proposes a new address family identifier (AFI) for the
purpose of disseminating the locator-identifier mappings among the
nodes participating in the ILA domain. For example, this extension
could be used to propagate the mappings from ILA hosts to ILA routers
and allow the routers to perform their function (see
[I-D.herbert-nvo3-ila] for definition of the "ILA router" functions).

2. BGP ILA AFI

This document introduces a new AFI known as a "Identifier-Locator
Addressing AFI" (ILA AFI) with the actual value to be assigned by
IANA. The purpose of this AFI is disseminating the mapping
information in the ILA domain, e.g. between ILA hosts and ILA
routers. This document defines the use of SAFI values of "1"
(unicast) and "2" (multicast) only.

3. Capability Advertisement

A BGP speaker that wishes to exchange ILA mapping information MUST
use the Multiprotocol Extensions Capability Code, as defined in
[RFC4760], to advertise the corresponding AFI/SAFI pair.

4. Disseminating Identifier-Locator mapping information

4.1. Advertising ILA mapping information

For the purpose of ILA mapping encoding, the 8-octet locator field
SHALL be encoded in the "next-hop address" field. The "length of the
next-hop address" MUST be set to "8". The identifiers bound to the
locator SHALL be encoded within the NLRI portion of MP_REACH_NLRI
attribute. The NLRI portion of MP_REACH_NLRI starts with the two-
octet "Length of identifiers" field. The rest of the NLRI is a
collection of 8-octet identifiers that are bound to the locator
specified in the "next-hop address" field.

```
+----------------------------------------------------------+
| Address Family Identifier (2 octets)                     |
+----------------------------------------------------------+
| Subsequent Address Family Identifier (1 octet)           |
+----------------------------------------------------------+
| Length of Next Hop Address (1 octet, set to "8")         |
+----------------------------------------------------------+
| Locator value (8 octets)                                 |
+----------------------------------------------------------+
| Reserved (1 octet), must be zero                         |
+----------------------------------------------------------+
| Length of NLRI field (2 octets, multiple of 8)           |
+----------------------------------------------------------+
| Identifiers (variable, 8 octets each)                    |
+----------------------------------------------------------+
```

Figure 1: MP_REACH_NLRI Layout

4.2. Withdrawing ILA mapping information

   Withdrawal of ILA mapping information is performed via an
   MP_UNREACH_NLRI attribute advertisement organized as following:

```
    +----------------------------------------------------------+
    | Address Family Identifier (2 octets)                     |
    +----------------------------------------------------------+
    | Subsequent Address Family Identifier (1 octet)           |
    +----------------------------------------------------------+
    | Length of witdrawn identifiers (2 octets)                |
    +----------------------------------------------------------+
    | Identifiers (variable, 8 octets each)                    |
    +----------------------------------------------------------+
```

                    Figure 2: MP_UNREACH_NLRI Layout

5. Interpreting the mapping information

5.1. Unicast SAFI

   Only the locator part of ILA address is used for packet routing, and
   every node that hosts an identifier MUST have a unique routable /64
   prefix within the scope of the ILA domain. The identifiers
   advertised under the ILA AFI are expected to be used by the data-
   plane implementation to perform match on a full IPv6 address and
   decide whether the locator portion of the address needs a re-write.
   It is up to the implementation to decide which full 128-bit IPv6
   addresses need a rewrite, e.g. by matching on a Standard Identifier
   Representation (SIR) prefix as defined in [I-D.herbert-nvo3-ila].

   The locator rewrite information comes from the next-hop "address"
   associated with the identifier. The next-hop field of MP_REACH_NLRI
   attribute MUST NOT be used for any routing resolutions/lookups by the
   BGP process itself. It should be used purely to create a rewrite
   rule in the data-plane forwarding table. The actual forwarding
   decision is then based on subsequent lookup in the forwarding table
   to find the next hop to send the packet to.

5.2. Multicast SAFI

   For multicast packets, the RPF check process SHALL be modified for
   use with ILA source addresses. Specifically, source ILA IPv6
   addresses with the identifier portion matching the mapping table
   SHALL be mapped to proper locator, prior to performing the RPF check.
   The ILA source addresses need to be identified by some means specific
   to ILA implementation, e.g. by matching on configured SIR prefixes.

The ILA addresses that do not match any mapping entry SHALL be
considered as failing the RPF check.

6. Inter-domain mapping exchange

The ILA mappings are only unique with an ILA domain - it is possible
that different domains may re-use the same identifiers. To make
identifiers globally unique, they MUST be concatenated with the SIR
prefix assigned to each ILA domain. These globally unique
identifiers may then be exchanged among the ILA domains. For the
purpose of facilitating this exchange, IANA will be requested to
allocate new SAFI value called "VPN-ILA" SAFI. The MP_REACH_NLRI
attributes exchanged over this SAFI will look as following:

```
+----------------------------------------------------------+
| Address Family Identifier (2 octets)                     |
+----------------------------------------------------------+
| Subsequent Address Family Identifier (1 octet)           |
+----------------------------------------------------------+
| SIR prefix (8 octets)                                    |
+----------------------------------------------------------+
| Length of Next Hop Address (1 octet, set to "8")         |
+----------------------------------------------------------+
| Locator value (8 octets)                                 |
+----------------------------------------------------------+
| Reserved (1 octet), must be zero                         |
+----------------------------------------------------------+
| Length of NLRI field (2 octets, multiple of 8)           |
+----------------------------------------------------------+
| Identifiers (variable, 8 octets each)                    |
+----------------------------------------------------------+
```

The main change from the Unicast/Multicast SAFI's is presence of the
SIR prefix field in the announcement. Correspondingly, the
MP_UNREACH_NLRI will look as following:

```
+----------------------------------------------------------+
| Address Family Identifier (2 octets)                     |
+----------------------------------------------------------+
| Subsequent Address Family Identifier (1 octet)           |
+----------------------------------------------------------+
| SIR Prefix (8 bytes)                                     |
+----------------------------------------------------------+
| Length of witdrawn identifiers (2 octets)                |
+----------------------------------------------------------+
| Identifiers (variable, 8 octets each)                    |
+----------------------------------------------------------+
```

The use of domain-specific identifiers would require the ILA hosts
and routers to make their ILA cache lookups based on the full 128-bit
prefix.

[7](#). Use of Add-Paths extension with ILA AFI

It could be useful to bind multiple locators to the same identifier,
e.g. for the purpose of load-sharing. To make announcing multiple
locators possible, the MP_REACH_NLRI and MP_REACH_NLRI attributes are
extended to encode the path-identifier per [[I-D.ietf-idr-add-paths](#)],
correspondingly enabling the capability for the AFI/SAFI.
Specifically, the NLRI field will look as following:

```
+----------------------------------------------------------+
| Path identifier (4 octets)                               |
+----------------------------------------------------------+
| Length of NLRI field (2 octets, multiple of 8)           |
+----------------------------------------------------------+
| Identifiers (variable, 8 octets each)                    |
+----------------------------------------------------------+
```

This encoding instructs the receiver to create multiple locator
entries for an identifier, differentiated by their path identifiers.
Optionally, a weight could be associated with each path using the
"link bandwidth" extended community defined in
[[I-D.ietf-idr-link-bandwidth](#)]

[8](#). IANA Considerations

For the purpose of this work, IANA would be asked to allocate a value
for the new AFI, and the VPN-ILA SAFI.

[9](#). Manageability Considerations

TBD

[10](#). Security Considerations

This document does not introduce any changes in terms of BGP
security. Defining ILA security model is outside of scope of this
document.

[11](#). Acknowledgements

The author would like to thank Doug Porter for the initial idea and
discussion of this proposal.

12. References

12.1. 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>.

   [I-D.ietf-idr-add-paths]
              Walton, D., Retana, A., Chen, E., and J. Scudder,
              "Advertisement of Multiple Paths in BGP", draft-ietf-idr-
              add-paths-13 (work in progress), December 2015.

12.2. Informative References

   [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>.

   [I-D.herbert-nvo3-ila]
              Herbert, T., "Identifier-locator addressing for network
              virtualization", draft-herbert-nvo3-ila-02 (work in
              progress), March 2016.

   [I-D.ietf-idr-link-bandwidth]
              Mohapatra, P. and R. Fernando, "BGP Link Bandwidth
              Extended Community", draft-ietf-idr-link-bandwidth-06
              (work in progress), January 2013.

Author's Address

   Petr Lapukhov
   Facebook
   1 Hacker Way
   Menlo Park, CA 94025
   US

   Email: petr@fb.com