           Use of BGP for dissemination of ILA mapping information
                    draft-lapukhov-bgp-ila-afi-02

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].

Table of Contents

1. Introduction

   Under the ILA proposal, 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).
Additional information that provides more detailed examples of
deployment scenarios using BGP could be found in
[I-D.lapukhov-ila-deployment].

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, with the value being multiple of
8. 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 is expected to have a unique /64
   prefix routable 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 in general is not expected to be used for any routing resolutions/lookups by the BGP process. It should primarily be used 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.

In some scenarios, resolving the next-hop attribute via additional lookups tables might be necessary. For example, for environments that deploy MPLS ([RFC3031]) forwarding, the locator may resolve to a label stack that is required to perform further forwarding. In this case, the ILA address rewrite will be accompanied by additional actions, such as label stack imposition. These decisions have to be made in implementation dependant fashion.

5.2. Multicast SAFI

Multicast SAFI retain the same encoding format, with a different SAFI value. 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 between multiple ILA domains. IANA will be requested to allocate new SAFI value called "VPN-ILA" SAFI to facilitate exchange of inter-domain ILA mappings. 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 difference 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. BGP Next-Hop attribute handling with ILA

This document proposes that the BGP next-hop attribute value encode
the locator associated with all identifiers found in MP_REACH_NLRI
attribute. It is possible that an intermediate speaker may change
the next-hop value. This may be required to ensure all traffic for
the associated identifiers is routed through that intermediate
speaker. The speaker is expected to maintain the original ILA
mappings in its mapping table, and perform additional destination

address translation for the ILA packets. This way, a form of loose
hop traffic engineering could be realized within an ILA domain.

It is common that BGP implementations reset the next-hop value for
announcements made over eBGP sessions. Such scenario may be common
between two different ILA domains.

8. 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)                    |
+----------------------------------------------------------+
```

Such 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](#)]

9. 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.

10. Manageability Considerations

The ILA mappings distribution is likely to be done using separate
infrastructure, independent from the BGP topology used for regular
routing information distribution. Most likely there will be a
collection of iBGP route-reflectors deployed within each ILA domain,
and peering with a BGP process running on each ILA router and ILA
host. More details and deployment examples could be found in
[[I-D.lapukhov-ila-deployment](#)].

11. 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.

12. Acknowledgements

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

13. References

13.1. 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>.

   [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
              Label Switching Architecture", RFC 3031,
              DOI 10.17487/RFC3031, January 2001,
              <http://www.rfc-editor.org/info/rfc3031>.

   [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-15 (work in progress), May 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.

13.2. 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 IPv6",
             draft-herbert-nvo3-ila-03 (work in progress), October
             2016.

   [I-D.lapukhov-ila-deployment]
             Lapukhov, P., "Deploying Identifier-Locator Addressing
             (ILA) in datacenter", draft-lapukhov-ila-deployment-00
             (work in progress), March 2016.

Author's Address

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

   Email: petr@fb.com