MP-BGP Extension and the Procedures for IPv4/IPv6 Mapping Advertisement
draft-xie-idr-mpbgp-extension-4map6

Chongfeng Xie
Guozhen Dong
China Telecom

Xing Li
Congxiao Bao
CERNET Center/Tsinghua University

Guoliang Han
Indirection Network Inc.

Davey Song
Zhongfeng Guo
Alibaba Cloud

idr@IETF 117, July 2023
Overview

• [draft-ietf-v6ops-framework-md-ipv6only-underlay] proposes a framework, in which IPv4 packets will be stateless translated or encapsulated into IPv6 ones for transmission across multi-domain IPv6-only underlay.

• This document defines MP-BGP extension and the procedures for IPv4 service delivery in multi-domain IPv6-only underlay networks.

• It was proposed in Jan. 2023, firstly presented in IETF 116, current version is -03.

• Existing approaches, e.g. RFC5565(Softwire Mesh Framework) and RFC6992 (Routing for IPv4-Embedded IPv6 Packets), are mainly aimed at a single domain.
**MP_REACH_NLRI**

- The existing AFI/SAFI combination is used to identify the reachability of IPv4 address block in IPv6-only network.
  - AFI = 2 (IPv6)
  - SAFI = 1 (Unicast)
  - Length of Next Hop
  - Network Address of Next Hop
  - NLRI: Composite IPv6 address prefix, which is composed of a **IPv6 mapping prefix, the original IPv4 address prefix**, and a series of padding bits

```
  | L2 |
IPv6 Mapping Prefix | IPv4 address prefix |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td></td>
</tr>
</tbody>
</table>
```

A series of zeros for padding
4map6 BGP path attribute

- It specifies the IPv6 mapping prefix and other additional information needed to properly transform the IPv4 packets.
- It is optional and transitive.

<table>
<thead>
<tr>
<th>Length of IPv6 Mapping Prefix (1 octets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forwarding Type (1 octet)</td>
</tr>
<tr>
<td>Address Origin Type (1 octet)</td>
</tr>
<tr>
<td>IPv4 Original ASN (4 octets)</td>
</tr>
</tbody>
</table>

Note: ATTR_SET attribute of RFC6368 can also be used to transfer the BGP routing information of the IPv4 in IPv6-only networks, including IPv4 ASN,
Revisions made since IETF 116

- Based on the comments of Jeff Hass, ATTR_SET attribute of RFC6368 can be used to transfer the BGP routing information of the IPv4 in IPv6-only networks, this has been added in section 3.2.
- The use case of IPv6-only DC for AI-fabric is added in the appendix.
- Several editorial changes have been made.
- Davey Song and Zhongfeng Guo of Alibaba Cloud have joined as co-authors.
System Implementation and Test

IPv4 network A

IPv4 Client

AS #1

IPv6

2001:da8:282:2402::/64

P1

AS #2

IPv6

2001:da8:282:2408::/96

AS #3

IPv6

2001:da8:282:2401::/96

IPv4 Internet

IPv4 Network B

Multi-domain IPv6-only network
Original IPv4 Route vs New IPv6 Route

IPv4 route in PE2

- BGP table version is 11, local router ID is 192.168.66.189, vrf id 0
- Default local pref 100, local AS 65489
- Status codes: s suppressed, d damped, h history, * valid, > best, = multipath, i internal, r RIB-failure, S Stale, R Removed
- NextHop: (none) nextHop's vrf id, < announce-nh-self
- Origin codes: i - IGP, e - EGP, ? - incomplete

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>110.242.68.0/24</td>
<td>0.0.0.0</td>
<td>0</td>
<td>32768</td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>192.168.66.0/32</td>
<td>0.0.0.0</td>
<td>0</td>
<td>32768</td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>192.168.66.173/32</td>
<td>192.168.66.173</td>
<td>0</td>
<td>64520</td>
<td>i</td>
<td></td>
</tr>
<tr>
<td>192.168.66.175/32</td>
<td>192.168.66.175</td>
<td>0</td>
<td>64550</td>
<td>i</td>
<td></td>
</tr>
</tbody>
</table>

Displayed 4 routes and 4 total paths

IPv6 route received by P router

- xlate# show bgp ipv6 2001:da8:282:2408::6ef2:4400/120
- BGP routing table entry for 2001:da8:282:2408::6ef2:4400/120
- Paths: (1 available, best #1, table default)
- Advertised to non-peer group peers:
  - 2001:da8:282:ff0c:9ec5:104:c664:171
  - 64589
- Origin IGP, metric 0, valid, external, best (First path received)
- Last update: Thu Jul 13 02:30:19 2023

110.242.68.0/24 is in IPv4 Internet
(110.242.68.3 is www.baidu.com)

IPv6 Mapping Prefix (/96)

- 2001:da8:282:2408:: 6e f2 : 44 00/120
- 110. 242. 68. 0/24
## 4map6 BGP Path Attribute

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0xc0</td>
<td>Path Attribute Flags (optional transitive)</td>
</tr>
<tr>
<td>0x40</td>
<td>Path Attribute Type Code (only for trial)</td>
</tr>
<tr>
<td>0x03</td>
<td>DATA length (3 bytes)</td>
</tr>
<tr>
<td>0x60</td>
<td>IPv6 Mapping Prefix Length (96)</td>
</tr>
<tr>
<td>0x02</td>
<td>Forwarding Type (translation)</td>
</tr>
<tr>
<td>0x00</td>
<td>Address Origin Type (local)</td>
</tr>
</tbody>
</table>
IPv6-IPv4 auto-config in PE1

PE1 received IPv6 route 2001:da8:282:2408::6ef2:4400/120 (attribute plen=96)

```
xlat# show bgp ipv6 2001:da8:282:2408::6ef2:4400/120
BGP routing table entry for 2001:da8:282:2408::6ef2:4400/120
Paths: (1 available, best #1, table default)
    Advertised to non peer-group peers:
    64589
    Origin IGP, metric 0, valid, external, best (First path received)
    Last update: Thu Jul 13 02:30:19 2023
```

PE1 map this route to an IPv4-IPv6 translation rule:

```
```

```
local IPv6 prefix: 2001:da8:282:2402::/64
local map6id: fec0::1/64
local mapid: 172.2.0.11/24
local nexthop: 2001:da8:282:ff0c:aa4f:44c9:84a0:170
local prefix6 address: 64:ff9b::/96
local dnat ns: dnat66
local dnat ipset name: iviDns64-ipset
```
IPv6 host → IPv4 server visit

Support IPv6 → IPv6, IPv4 → IPv4, IPv6 → IPv4, etc.

ping IPv4 Internet server from IPv6 hosts

windows IPv6 host

MacOS IPv6 host

Android IPv6 host
Next Steps

• Comments and suggestions are welcome, and make further refinement to improve the document

• Authors would like to ask for WG adoption of this document
Thank you!

Q&A