Workgroup: Network Working Group
Internet-Draft:
draft-xie-idr-mpbgp-extention-4map6-00
Published: 15 January 2023
Intended Status: Standards Track
Expires: 19 July 2023
Authors: C. Xie
China Telecom
G. Dong
China Telecom
X. Li
CERNET Center/Tsinghua University

MP-BGP Extension and the Procedures for IPv4/IPv6 Mapping Advertisement

Abstract

This document defines NLRI with specific AFI/SAFI combination, a new BGP path attribute known as the "4map6" and a set of related procedures, which can be used for transferring IPv4/IPv6 address mapping rule to support IPv4 service delivery in multi-domain IPv6-only underlay networks.

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 https://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 19 July 2023.

Table of Contents

1. Introduction

The document [I-D./draft-ietf-v6ops-framework-md-ipv6only-underlay] proposes a framework for deploying IPv6-only as the underlay in multi-domain networks, in which IPv4 packets will be stateless translated or encapsulated into IPv6 ones for transmission across IPv6-only underlay domains. To achieve this goal, this framework introduces a specific data structure called address mapping rule to support stateless IPv4-IPv6 packet conversion. For an incoming packet, the mapping rules are used by the ingress PE to generate corresponding IPv6 source and destination addresses from the IPv4 source and destination address of the original IPv4 packet, and vice versa. Since the mapping rule for the destination IPv4 address can identify the right PE egress by providing the IPv6 mapping prefix, it gives the direction of IPv4 service data transmission throughout the IPv6-only network. It is obvious that the exchange of the mapping rule corresponding to the destination IPv4 address in a given packet should precede to the process of IPv4 data transmission in IPv6-only network, otherwise, the data originated from IPv4 network will be dropped due to the absence of the IPv6 mapping prefix corresponding to its destination address.

When an ingress PE processes the incoming IPv4 packets, the mapping rule for the source address can be obtained locally, but for the

mapping rule of the destination address, since it is not generated locally by the ingress PE, it needs corresponding methods to be obtained remotely. This document defines MP-BGP extension in which BGP update message contains the mapping rule for IPv4 service delivery. The extensions include NLRI with specific AFI/SAFI combination, new BGP Path Attribute known as the "4map6" corresponding to the NLRI and a set of related procedures.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14[RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Terminology and Reference Topology

In the context of this document, multi-domain underlay networks refer to a network system composed of multiple autonomous systems (i.e., AS) interconnected, each AS can serve different scenarios. Multi-domain networks can be operated by one or more network operators. Consider the following scenarios, the network shown in figure 1 is typical multi-domain IPv6-only underlay networks, it is used as a basic scenario to illustrate the extension of the MP-BGP and its related procedures in this document. The network comprises of AS1, AS2 and AS3, it provides IPv4 services communications between IPv4 network 1 and IPv4 network 2, which have IPv4 address block IPv4Blk1 and IPv4Blk2 respectively. It is consistent with section 6 of draft [I-D.ietf-v6ops-framework-md-ipv6only-underlay] . PE and P routers are network routers which constitute the IPv6-only underlay. The definition of PE and P is consistent with that in draft [I-D.ietf-v6ops-framework-md-ipv6only-underlay] . It should be noted that in a multi-domain networks, some ASBRs are not at the edge of the network. In this case, they run as P routers. On each PE router that the IPv4 address prefix is reachable through, there is a locally configured IPv6 virtual interface (VIF) address. The VIF address, as an ordinary global IPv6 /128 address, must also be injected into the IPv6 IGP so that it is reachable across the multi-domain transit core.

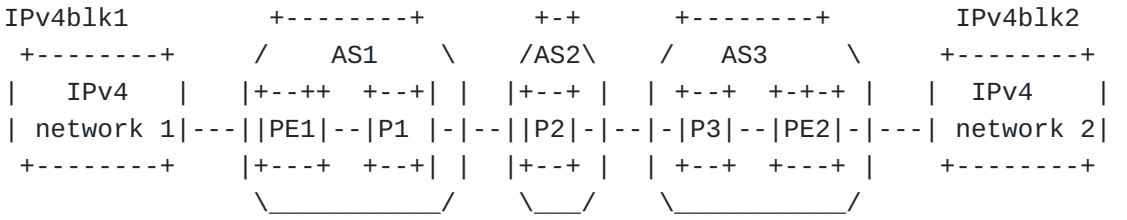

Figure 1.Topology of Typical Multi-domain IPv6-only Networks

The following term will be used in this document,

o Distance metric, the distance to the egress PE in terms of the number of ASes.

The extension of MP-BGP4 for mapping rule processing and transmission across domains in this draft will involve PE and P routers. Each PE and P router maintains a Mapping Rule Database as depicted in figure 2. The entry in the Mapping Rule Database consists of an IPv4 address prefix, IPv4 address prefix length, IPv6 mapping prefix, IPv6 mapping prefix length and the distance to the egress. It should be noted that the database here is just an example, and developers can design the structure of database according to the actual situation.

IPv4 Address Prefix	IPv4 Address Prefix Length	IPv6 Mapping Prefix	IPv6 Mapping Prefix Length	Distance to the Egress

Figure 2: The Entry of Mapping Rule Database

The IPv4 packet sent from IPv4 network 1 will traverse the IPv6-only network and reach the destination network, i.e., IPv4 network 2. Its ingress in the IPv6-only network is PE1 and the egress is PE2. Before the data packet is transmitted, the address mapping rules corresponding to its IPv4 destination address should be transmitted from PE2 to PE1. During the mapping rule announcement and transmission process, it may pass through the intermediate nodes, such as P3, P2 and P1, and finally reach PE1. For a given intermediate node, it may receive advertisement messages of this mapping rule from multiple upstream intermediate nodes. In order to reduce the overall quantity of advertisement message, it needs to select and update the local Mapping Rule Database, generates advertisement messages based on the selected mapping rule information and transmit them to downstream intermediate nodes.

3. MP-BGP Protocol Extension

3.1. NLRI Encoding for IPv4/IPv6 Mapping Rule Advertisement

Multiprotocol BGP (MP-BGP) [RFC4760]specifies that the set of usable next-hop address families is determined by the Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI). [RFC8950] specifies the extensions to allow advertisement of IPv4 NLRI or VPN IPv4 NLRI with a next-hop address that belongs to the IPv6 protocol. This document specifies the extensions necessary to allow advertisement of address mapping rules across domains. To support the transmission of mapping rule from any egress PE to any

ingress PE within and across domains, a new Subsequent Address Family Identifier (SAFI) needs to be assigned by IANA. When this is available, the BGP update message can contain the 4map6 BGP path attribute. The BGP update whose MP_REACH_NLRI attribute contains the AFI/SAFI combinations specified above is called as 4map6 routing information. The use and meaning of the fields of MP_REACH_NLRI in this case are as follows:

- AFI = 2 (IPv6)

- SAFI = xx (4map6)

- Length of Next Hop

- Network Address of Next Hop = When a BGP speaker advertises the 4map6 NLRI via BGP, it uses its own address as the BGP next hop in the MP_REACH_NLRI.

- NLRI = Composite IPv6 address prefix, which is composed of a IPv6 mapping prefix, the original IPv4 address prefix, and the remaining bits are zero.

The NLRI field is encoded as shown in figure 3:

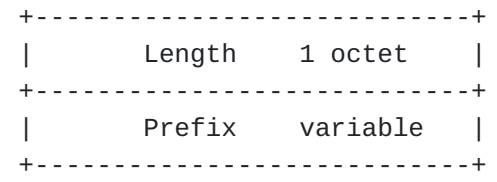

Figure 3: the Format of NLRI Field

3.2. 4map6 Path Attribute

This document specifies a way in which BGP protocol can be used by a given PE to tell other PE, "If you need to send IPv4 packet whose destination address is within a given IPv4 address block, please send them to me, here's the information you need to properly transform the IPv4 packet into IPv6 one". A PE signals this information to other BGP speakers by using a new BGP attribute type value -- the 4map6 attribute. This attribute specifies the IPv6 mapping prefix that may be used, as well as whatever additional information (if any) is needed in order to properly transform the IPv4 packets. As a new BGP path attribute defined in this document, 4map6 attribute is optional and transitive, it is encoded as shown below:

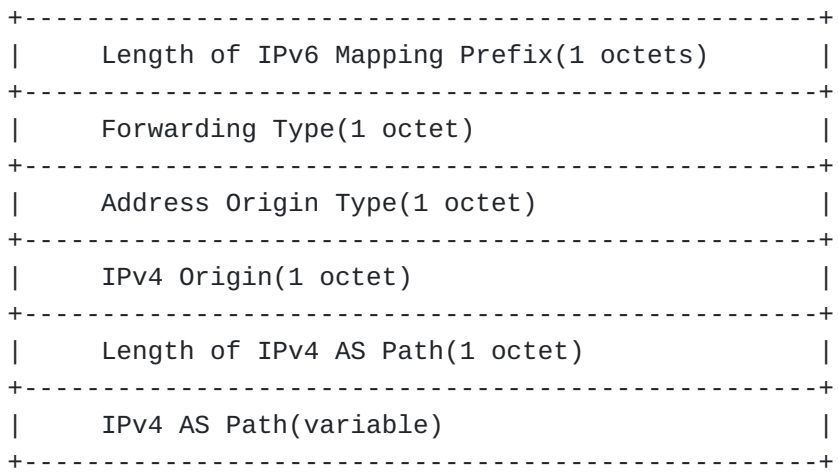

Figure 4:Encoding of the 4map6 attribute

The use and meaning of these fields are as follows:

a) Length of IPv6 Mapping Prefix

This is a 1-octet field whose value expresses the length of IPv6 mapping prefix.

b) Forwarding Type

This field identifies the IPv4/IPv6 forwarding capability of the egress PE, the data octet can assume the following values:

Value Meaning

0 Translation and encapsulation

1 Encapsulation

2 Translation

c) Address Origin Type

The data octet can assume the following value:

Value Meaning

0 Local

1 Relay

d) IPv4 Origin

This field is the copy of the Origin attribute in BGP update message received from IPv4 domain.The value of this field exists only when the value of "Address Origin Type" is 1, otherwise it is NULL.

e) Length of IPv4 AS_Path

A 1-octet field whose value expresses the length of the "IPv4 AS_Path" measured in octets.The value of this field exists only when the value of "Address Origin Type" is 1, otherwise it is NULL.

f) IPv4 AS_Path

This field is the copy of the AS_PATH attribute in BGP UPDATE message received from IPv4 domain.The value of this field exists only when the value of "Address Origin Type" is 1, otherwise it is NULL.

3.3. Explicit Withdrawal of IPv4/IPv6 Mapping Rule

When a PE ceases to provide egress service for a given IPv4 address block, it may explicitly withdraw the mapping rules associated with it. Suppose a PE has announced, on a given BGP session, the mapping rule of a given IPv4 prefix and it now wishes to withdraw that mapping rule. To do so, it may send a BGP UPDATE message with an MP_UNREACH_NLRI attribute.

This encoding of MP_UNREACH_NLRI attribute is used for explicitly withdrawing the mapping rule for a given IPv4 prefix (on a given BGP session). Note that IPv4 address prefix/IPv6 mapping prefix bindings that were not advertised on the given session can not be withdrawn by this method.

When using an MP_UNREACH_NLRI attribute to withdraw a IPv4 route whose NLRI was previously specified in an MP_REACH_NLRI attribute, the lengths and values of the respective prefixes must match, and the respective AFI/SAFIs must match. An explicit withdrawal in an AFI/SAFI-xx UPDATE on a given BGP session not only withdraws the binding between the IPv4 address prefix and the IPv6 mapping prefix, it also withdraws the path to that prefix that was previously advertised in a SAFI-xx UPDATE on that session.

4. Operation

4.1. Advertisement of Mapping Rule Update by egress PE

When a PE router learns routing information from the locally attached IPv4 access networks, the control plane of the PE should process the information as follows:

1. Install and maintain local IPv4 routing information in the IPv4 routing database.

2. Install and maintain new entries in the Mapping Rule Database. Each entry should consist of the IPv4 prefix and the local IPv6 mapping prefix.

3. Advertise the new contents of the local Mapping Rule Database in the form of BGP update advertisement to IPv6 peer routers. The process to generate IPv6 route advertisement with 4map6 attribute based on IPv4 route advertisement messages is as follows:

a) Set the values of AFI and SAFI in MP_REACH_NLRI to 2 and xx respectively;

b) The IPv6 mapping prefix of the egress PE splices IPv4 address blocks in IPv4 routing advertisements to form a composite IPv6 address prefix with a length of L1. The composite IPv6 address prefix is copied to address prefix field of the NLRI structure in the MP_ REACH_NLRI, and the length field of the NLRI is set to L1, the structure of the composite IPv6 address prefix in NLRI is shown in figure 5. L2 is the length of the IPv6 mapping prefix Pref6-2 of PE2, the field of Length of IPv6 Mapping Prefix value in the 4map6 attribute is set to L2.

c) In addition, the values of Origin, Length of AS_ Path, AS_Path information in the original IPv4 route advertisement is copied to the fields of IPv4 Origin, Length of IPv4 AS_Path,IPv4 AS_Path of 4map6 attribute respectively.

Figure 5:Structure of IPv6 prefix in NLRI

4.2. Receiving Mapping Rule advertisement by P router

When a P router receives BGP update advertisement from neighboring P or PE routers and uses that information to populate the local Mapping Rule Database, the following procedures are used to update the Mapping Rule Database and send mapping rule advertisement to next equipment:

1. Validate the received BGP update advertisement as 4map6 routing information by AFI = 2 (IPv6) and SAFI = xx (4map6).

2. Extract the IPv4 address prefix which is encoded in positions L2 to L1-1 of the NLRI field and lookup in the Mapping Rule Database, if an entry which matches the IPv4 address prefix is found, then,

- Compare the distance metric in the 4map6 attribute of BGP advertisement and that of the entry found, if the former is less than the latter, then

    o Update the entry found in the Mapping Rule Database with the attributes of BGP advertisement by extracting the IPv6 address prefix from the IPv6 mapping prefix field and place that as an associated entry next to the IPv4 network index.

    o Advertise the updated contents of the local Mapping Rule Database in the form of MP_REACH_NLRI update information to IPv6 peer routers.

  else then

    o Keep the entry in the Mapping Rule Database unchanged.

    o Advertise the contents of the local Mapping Rule Database in the form of BGP update advertisement to IPv6 peer routers.

else then

- Install and maintain a new entry in the Mapping Rule Database with the extracted IPv4 prefix, its corresponding IPv6 mapping prefix and distance metric to the egress.

- Advertise the contents of the local Mapping Rule Database in the form of BGP update advertisement to IPv6 peer routers.

It should be noted that this process does not change or affect the IPv6 FIB table of the P router.

4.3. Receiving Mapping Rule Update by Ingress PE

When a PE router receives BGP advertisement from neighboring P or PE routers and uses that information to populate the local Mapping Rule Database and the BGP routing database, the following procedures are used to update the Mapping Rule Database and send IPv4 routing information to its IPv4 peers.

1. Validate the received BGP update advertisement as 4map6 routing information by AFI = 2 (IPv6) and SAFI = xx (4map6).

2. Extract the IPv4 address prefix which is encoded in positions L2 to L1-1 of the NLRI field and lookup in the Mapping Rule Database, if an entry which matches the IPv4 address prefix is found, then,

- Compare the distance metric in the BGP advertisement and that of the entry found, if the former is less than the latter, then

o Update the entry found in the Mapping Rule Database with the 4map6 attributes of BGP advertisement by extracting the IPv6 address prefix from the IPv6 mapping prefix field and place that as an associated entry next to the IPv4 network index.

o Redistribute the new 4map6 routing information to the local IPv4 routing table. Set the destination network prefix as the extracted IPv4 prefix, set the Next Hop as Null, and set the OUTPUT Interface as the 4map6 VIF on the local PE router.

else then

o Keep the entry in the Mapping Rule Database unchanged.

else then

- Install and maintain a new entry in the Mapping Rule Database with the extracted IPv4 prefix, its corresponding IPv6 mapping prefix and distance metric to the egress.

- Redistribute the new 4map6 routing information to the local IPv4 routing table. Set the destination network prefix as the extracted IPv4 prefix, set the Next Hop as Null, and set the OUTPUT Interface as the 4map6 VIF on the local PE router.

5. Mapping Rule Capability

[RFC5492]defines a Capabilities Optional Parameter and processing rules. The Capabilities Optional Parameter is a triple that includes a one-octet Capability Code, a one-octet Capability length, and a variable-length Capability Value. A BGP speaker can include a Capabilities Optional Parameter to communicate capabilities in a BGP OPEN message. A PE or P router that wishes to exchange mapping rule information must use the Multiprotocol Extensions Capability Code as defined in [RFC4760], to advertise the corresponding (AFI, SAFI) pair.

6. Error Handling

When a BGP speaker encounters an error while parsing the 4map6 path attribute, the speaker must treat the update as a withdrawal of existing routes to the included 4map6 SAFI NLRIs, or discard the update if no such routes exist. A log entry should be raised for local analysis.

7. IANA Considerations

With this document IANA is requested to allocate the following codes,

1)A code for 4map6 path attribute in the BGP "BGP Path Attributes" registry

2)Value xx for 4map6 in the BGP "Capability Codes" registry

3)A new SAFI value (xx) for the BGP 4map6 SAFI.

All the codes above use this document as the reference.

8. Security Considerations

This extension to MP-BGP does not change the underlying security issues inherent in the existing MP-BGP.

9. References

9.1. 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, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <https://www.rfc-editor.org/info/rfc4760>.

[RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <https://www.rfc-editor.org/info/rfc5492>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8950] Litkowski, S., Agrawal, S., Ananthamurthy, K., and K. Patel, "Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop", RFC 8950, DOI 10.17487/RFC8950, November 2020, <https://www.rfc-editor.org/info/rfc8950>.

9.2. Informative References

[I-D.ietf-v6ops-framework-md-ipv6only-underlay]
Xie, C., Ma, C., Li, X., Mishra, G. S., Boucadair, M., and T. Graf, "Framework of Multi-domain IPv6-only Underlay Networks and IPv4-as-a-Service", Work in Progress, Internet-Draft, draft-ietf-v6ops-framework-md-ipv6only-underlay-00, 9 January 2023, <https://

www.ietf.org/archive/id/draft-ietf-v6ops-framework-md-ipv6only-underlay-00.txt>.

Authors' Addresses

Chongfeng Xie
China Telecom
Beiqijia Town, Changping District
Beijing
102209
China

Email: xiechf@chinatelecom.cn

Guozhen
China Telecom
Beiqijia Town, Changping District
Beijing
102209
China

Email: donggz@chinatelecom.cn

Xing Li
CERNET Center/Tsinghua University
Shuangqing Road No.30, Haidian District
Beijing
100084
China

Email: xing@cernet.edu.cn