Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 7 June 2026

C. Xie
G. Dong
China Telecom
X. Li
CERNET Center/Tsinghua University
G. Han
Indirection Network Inc.
Z. Guo
Alibaba Cloud

4 December 2025

MP-BGP Extension and the Procedures for IPv4/IPv6 Mapping Advertisement
draft-ietf-idr-mpbgp-extension-4map6-05

Abstract

This document defines MP-BGP extension and the procedures for IPv4 service delivery in multi-domain IPv6-only underlay networks. It defines a new TLV in the BGP Tunnel Encapsulation attribute, used in conjunction with a specific AFI/SAFI combination for advertising IPv4-over-IPv6 mapping rules. The behaviors of each type of network (IPv4 and IPv6) are also illustrated. In addition, this document provides the deployment and operation considerations when the extension is deployed.

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 7 June 2026.

Copyright Notice

Table of Contents

1. Introduction

The document [I-D.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 IPv4/IPv6 address mapping rule to support stateless IPv4-IPv6 packet conversion at the edge of the network. For brevity, in the rest of the document, we will refer to the IPv4/IPv6 address mapping rule as mapping rule. For an incoming IPv4 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 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 a new TLV for 4map6 tunnel type in BGP Encapsulation attribute corresponding to specific AFI/SAFI combination and related procedures in IPv6-only networks.

It should be noted that the approach in this document focuses on controlled environment, for instance, one network of single network operator or small network of cooperating network operators. One effect of using this approach is that the IPv4 address prefix in it will be given one IPv6 mapping prefix and advertised in a IPv6 mapping route.

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] and [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Reference Topology

In the context of this document, multi-domain underlay network refers to a network system composed of multiple autonomous systems (i.e., AS) interconnected, each AS can serve different scenarios. Multi-domain network can be operated by one or more network operators. Consider the following scenarios, the network shown in figure 1 is a typical multi-domain IPv6-only underlay network, it is used as a basic scenario to illustrate the extension of the MP-BGP and its related procedures in this document. The whole network comprises of AS1, AS2 and AS3, it provides IPv4 services communications between IPv4 network N1 and IPv4 network N2, which have IPv4 address block IPv4 A1 and A2 respectively. It is consistent with section 6 of

draft [I-D.ietf-v6ops-framework-md-ipv6only-underlay].

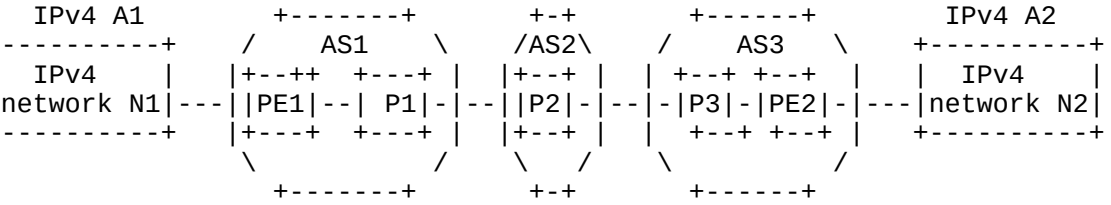

Figure 1:Topology of Typical Multi-domain IPv6-only Network

PE and P routers are network devices 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 multi-domain network, 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.

The extension of MP-BGP for mapping rule processing and transmission across domains in this document will involve PE routers, which setup BGP sessions to directly exchange 4map6 routing without affecting the RIB and FIB of intermediate P devices. Each PE router maintains a Mapping rule Database (i.e., MD) 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 of the PE, IPv6 mapping prefix length, address origin type and forwarding Type of the PE. 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	Address Origin Type	Forwarding Type

Figure 2: Entry structure of Mapping Rule Database

The IPv4 packet sent from IPv4 network N1 will traverse the IPv6-only network and reach the destination network, i.e., IPv4 network N2. Its source address and destination address are within IPv4 address block A1 and A2 respectively. Its ingress in the IPv6-only network

is PE1 and the egress is PE2. Before the data packet is transmitted, the mapping rules corresponding to IPv4 address block A2 should be transmitted from PE2 to PE1.

This mechanism is also in line with the requirements of emerging scenarios such as DCN for AI infra fabric, as described in Appendix A.

3. MP-BGP Protocol Extension

3.1. NLRI Encoding for Mapping Rule Advertisement

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 packets into IPv6 ones". 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 support the transmission of mapping rule from any egress PE to any ingress PE within and across domains. Since it is based on IPv6-only routing paradigm, it leverages the combination of AFI and SAFI, with the value of 2 (IPv6) and a to-be-assigned SAFI value (4map6) respectively, which identifies NLRI used for IPv4 forwarding in IPv6 network. In addition, to support the transmission of additional information of the mapping rule, it defines a new TLV for the 4map6 Tunnel Type in the BGP Tunnel Encapsulation attribute. With this new TLV, Address Origin Type and Forwarding Type can be obtained from BGP update by the ingress PE to properly transform the IPv4 packets. The route carried in the BGP update whose MP_REACH_NLRI attribute contains the AFI/SAFI combinations and the 4map6 TLV specified above is referred to as an ?IPv6 mapping route?.

The use and meaning of the fields of MP_REACH_NLRI in this case are as follows:

AFI = 2 (IPv6)

SAFI = xxx (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 = Synthetic IPv6 address prefix, which is composed of a IPv6 mapping prefix, the original IPv4 address prefix, and their length. Following section 5 of [RFC4760], the corresponding NLRI field for IPv4/IPv6 address mapping is encoded as shown in figure 3:

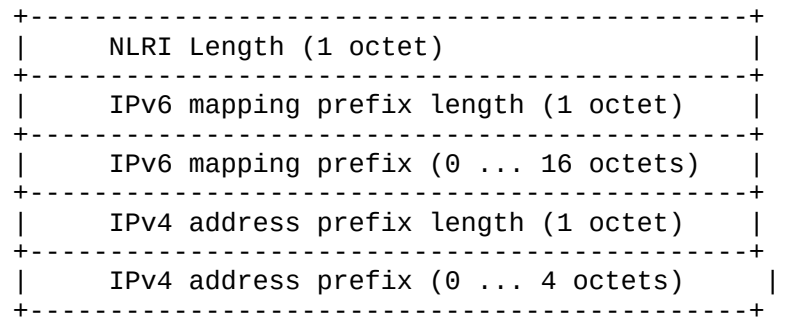

Figure 3:Format of NLRI Field

3.2. 4map6 Tunnel TLV

[RFC9012]defines a BGP path attribute known as the "Tunnel Encapsulation attribute", which can be used with BGP UPDATEs of various Subsequent Address Family Identifiers (SAFIs) to provide information needed to create tunnels and their corresponding encapsulation headers. It provides encodings for a number of tunnel types, along with procedures for choosing between alternate tunnels and routing packets into tunnels. BGP path attribute is composed of a set of Type-Length-Value (TLV) encodings. Each TLV contains information corresponding to a particular tunnel type. Since IPv4 data forwarding in [I-D.ietf-v6ops-framework-md-ipv6only-underlay] adopts stateless IPv4/IPv6 address mapping to generate IPv6 addresses, and supports both encapsulation and translation, a new tunnel type named 4map6 is defined in this document. Following the format of Tunnel Encapsulation TLV defined in section 2 of [RFC9012], the TLV for the 4map6 Tunnel is shown as follows,

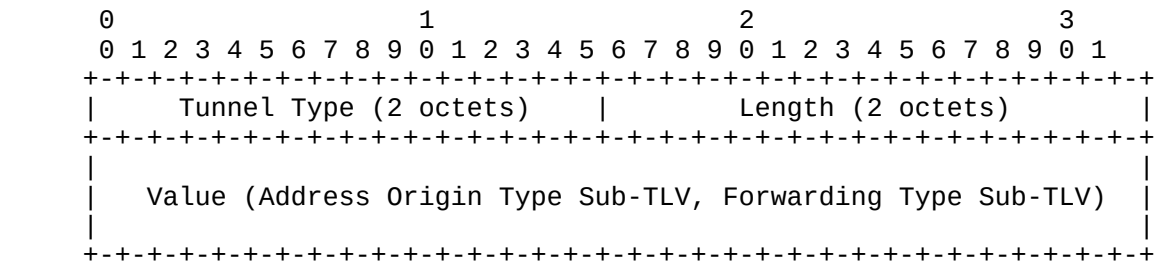

Figure 4:Format of the 4map6 Tunnel TLV

The code of "Tunnel Type" for the 4map6 Tunnel needs to be applied from the IANA registry "BGP Tunnel Encapsulation Attribute Tunnel Types" [IANA-BGP-TUNNEL-ENCAP].

The Value field contains Sub-TLVs. Per [RFC9012], each Sub-TLV consists of three fields: A 1-octet type, a 1-octet or 2-octet length (depending on the type), and zero or more octets of value. At present, two sub-TLVs are defined for the 4map6 Tunnel TLV: "Address Origin Type" and "Forwarding Type", they are arranged in sequential order in the Value field. These two Sub-TLVs are specified as follows:

a) The Address Origin Type Sub-TLV (Type Code yy1)

The Address Origin Type Sub-TLV specifies the type of the origin of IPv4 address prefix. The code of "Sub-TLV Type" for the Address Origin Type needs to be applied from the IANA registry "BGP Tunnel Encapsulation Attribute Sub-TLVs" [IANA-BGP-TUNNEL-ENCAP]. Its Value field can assume the following values:

Value Meaning

0 The IPv4 address prefix originates from the local configuration of the egress PE.

1 The IPv4 address prefix is obtained by egress PE from external IPv4 networks.

b) The Forwarding Type Sub-TLV (Type Code yy2)

The Forwarding Type Sub-TLV identifies the IPv4/IPv6 forwarding capability of the egress PE. The code of "Sub-TLV Type" for the Forwarding Type needs to be applied from the IANA registry "BGP Tunnel Encapsulation Attribute Sub-TLVs" [IANA-BGP-TUNNEL-ENCAP]. Its Value field it can assume the following values:

Value Meaning

0 Translation and encapsulation

1 Encapsulation

2 Translation

As a new type of Tunnel TLV, 4map6 Tunnel follows the validation rules of [RFC9012]. Per section 13 of RFC9012, the final octet of a 4map6 Tunnel TLV MUST also be the final octet of its final sub-TLV, i.e. the Address Origin Type Sub-TLV. If this is not the case, the 4map6 TLV MUST be considered to be malformed, and the "Treat-as-withdraw" procedure of [RFC7606] of is applied.

Furthermore, If a Tunnel Encapsulation attribute can be parsed correctly but contains a 4map6 Tunnel TLV whose tunnel type is not recognized by a particular BGP speaker, that BGP speaker MUST NOT consider the attribute to be malformed. Rather, it MUST interpret the attribute as if that 4map6 Tunnel TLV had not been present. If the route carrying the Tunnel Encapsulation attribute is propagated with the attribute, the 4map6 Tunnel TLV MUST remain in the attribute.

In addition, ATTR_ SET attribute(type code 128), defined in [RFC 6368], can be used to transfer the routing information of the IPv4 network in multi-domain IPv6-only network.

4. Operation

4.1. Advertisement of Mapping Rule Update by Egress PE

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

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

2. Generate BGP update advertisement based on the IPv4 route advertisement and the capability of the egress PE as follows:

    1) Set the values of AFI and SAFI in MP_REACH_NLRI to 2 and xxx respectively.

2) Following the NLRI coding method in section 3.1, generate a new NLRI using IPv6 mapping prefix length, IPv6 mapping prefix of the egress PE, IPv4 address prefix length and IPv4 address prefix.

3) Generate the 4map6 Tunnel TLV based on Forwarding Type of the egress PE and Address Origin Type of the IPv4 address prefix.

4) The Origin ASN, Length of AS_Path, AS_Path in the original IPv4 route advertisement are copied to the corresponding fields of ATTR_ SET attribute respectively.

3. Send the BGP update advertisement generated to its IPv6 peer routers.

4.2. Receiving Mapping Rule Update by Ingress PE

When a PE router receives BGP update advertisement from other 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 IPv4 routing information to its IPv4 peers.

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

2. Extract the IPv6 Mapping Prefix which is encoded in the NLRI field and check whether it is reachable in the IPv6 underlay network, that is, if a routing entry containing it can be found in the underlying IPv6 routing table. If yes, then proceed to the next step. Otherwise, this indicates that it is unreachable, then do not proceed to the next step.

3. Extract the IPv4 address prefix which is encoded in the NLRI field and lookup in Mapping rule Database, if an entry which matches the IPv4 address prefix is found, then,

1) Update the entry found in the Mapping rule Database with the IPv6 mapping prefix, Address Origin Type and Forwarding Type extracted from BGP advertisement, then place that as an associated entry next to the IPv4 network index.

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

else then

1) Install and maintain a new entry in the Mapping rule Database with the extracted IPv4 address prefix and its corresponding IPv6 mapping prefix, Forwarding Type, Address Origin Type.

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

As mentioned in [I-D.draft-ietf-v6ops-framework-md-ipv6only-underlay], multi-domain IPv6-only network supports both translation and encapsulation technologies for IPv4 data delivery at the forwarding layer. Take the encapsulation as an example, the reachability to the egress endpoint of tunnel may change over time, directly impacting the feasibility of the IPv4 service delivery. A tunnel that is not feasible at some moment may become feasible at later time when its egress endpoint address is reachable. The router may start using the newly feasible tunnel instead of an existing one. This may happen for translation-based data-path as well. How this decision is made is outside the scope of this document.

5. Error Handling

When a BGP speaker encounters an error while parsing the 4map6-related attributes, the speaker must treat the update as a withdrawal of existing IPv6 mapping routes, or discard the update if no such routes exist. A log entry should be generated for local analysis.

6. Deployment and Operation Considerations

6.1. Scalability Consideration

When operators use the new extension in actual IPv6 networks, it is necessary to consider its impact on BGP scalability. If there is not specific policy consideration during deployment, for the same IPv4 address block, different operators may use different prefixes to map it, so multiple synthesized IPv6 prefixes will be generated, which can have a significant impact on the scale of RIB and even FIB. Therefore, it is recommended that only one IPv6 mapping prefix should be configured for each IPv4 address block in principle, and this is also true in multi-operator scenario. In essence, the scalability issue is related to the strategy of IPv6 mapping prefix allocation. In section 6.3 of [I-D.ietf-v6ops-framework-md-ipv6only-underlay], two configuration mapping prefix methods are proposed, WKP (i. e., Well-Known Prefix) and NSP (i. e., Network Specific Prefix), which can be combined to assist in solving the scale concern. For different types of PE, it is recommended to use the following IPv6

mapping prefix configuration policy:

1) PE for access (Type 1), this kind of PE is configured with IPv4 address blocks, which are owned by the operator. Generally, it does not have direct interconnection with external IPv4 Internet. It is recommended to assigns NSP as the IPv6 mapping prefix to these address blocks and sets the sub-field of Address Origin Type in the 4map6 tunnel TLV to 0. Since the NSP is part of the operator's IPv6 address block, it is usually not necessary to advertise a dedicated IPv6 route for each IPv6 mapping prefix, so that no additional entries are added to the FIB of the P devices.

2) Interworking PE (Type 2), this kind of PE interworks directly with the external IPv4 Internet. For the address block in the IPv4 route announcement received from the IPv4 Internet, it is proposed to use WKP to configure the IPv6 mapping prefix, and set the sub-field of Address Origin Type in the 4map6 tunnel TLV to 1. With this configuration, in a IPv6 network with multiple interworking PE, regardless of which PE maps and advertisements the IPv6 mapped route, the NLRI for the received external IPv4 address block is the same, so there will be no significant impact on the scale of IPv6 network routing.

In addition, implementation of 4map6 related settings and policies at the network edge can also be useful to ensure scalability. For example, setting a policy on interworking PE to prohibit self-owned IPv4 address blocks from backtracking to IPv6 routing. Interworking PEs may receive BGP advertisement of self-owned IPv4 address blocks from IPv4 Internet. If a new IPv6 mapping route is generated using the mapping prefix of Interworking PEs and advertised in the IPv6 network, it will increase the load of RIB of routers and may form a routing loop. Therefore, it is necessary to configure policies on the PE to restrict the backtracking of IPv4 address blocks to be advertised in the IPv6 network. The strategy is also effective in multi-operator scenario. Moreover, Restricting the advertisement of BGP messages with Address Origin Type value of 1 to other operators can also help avoid situations that multiple operators assign different mapping prefixes to the same IPv4 address block.

Nevertheless, in some cases, such as when high reliability is required, some IPv4 address blocks need to be configured with multiple IPv6 mapping prefixes.

6.2. Route Distribution Control

With the approach in this document, IPv6 mapping routes should only be used within the closed multi-domain IPv6 network. To prevent IPv6 mapping routes from leaving the IPv6-only network and entering the IPv6 Internet, it is recommended to mark them when generating IPv6 mapping routes at the egress PE and set policies on the receiving PE to prevent leakage into the IPv6 Internet. For operators, the range of IPv6 mapping routes distribution can be controlled based on the new 4map6 SAFI, to be specific, BGP message with SAFI xxx is restricted from being advertised on the IPv6 Internet. In addition, as the IPv6 prefix of NLRI generated through mapping is longer than that of regular IPv6 routes, the sum of the lengths of IP mapping prefix and IPv4 address prefix is generally greater than 64. But in BGP routing between operators, there is a convention that the prefix of NLRI is required to be less than a certain length, such as 48 bits. If a route has a long prefix without 4map6 attribute, the receiving BGP router can filter it out. Furthermore, since the NLRI of the mapping route is synthesized based on legal IPv4 addresses and does not overlap with NLRIs of other native IPv6 routes, even if it is leaked to the IPv6 Internet, no traffic hijacking effect will occur.

7. Security Considerations

In the early stage of deployment, it can be expected that 4map6 extension is mainly used in small multi-domain IPv6 network with a few operator interconnections. At this stage, BGP collaboration was established on the basis of mutual trust between operators. In case of accidents or malfunctions, both parties can resolve them through collaborative means. Under this premise, when one egress PE of the other operator sends a 4map6 BGP announcement with Address Origin Type value of 0, the Ingress PE can trust the information in it, extract the items of the IPv4 address prefix, IPv6 mapping prefix, address origin type and forwarding type, then store them in the local Mapping rule Database. However, in the distant future, if the scope of 4map6 usage is further expanded, a dedicated authentication mechanism will be needed to verify the authenticity of 4map6 information in BGP advertisements, preventing malicious network operators from using their own address prefix to map other operators' IPv4 address blocks, thereby turning into network hijacking behavior, as stated in section 8.2 of [I-D.ietf-v6ops-framework-md-ipv6only-underlay], " The ability to advertise a mapping rule adds a new means by which an attacker could cause traffic to be diverted from its normal path." In this case, similar techniques as RPKI origin validation or IRR filtering are needed to help prevent this. Applying such technologies to the proposed mapping mechanism would mean that BGP prefix policy would

need to be able to be applied to the embedded IPv4 networks, this security enhancement can be defined in other documents.

8. IANA Considerations

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

1) A new SAFI value (xxx) for the BGP 4map6 in "Subsequent Address Family Identifiers (SAFI) Parameters" registry.

2) A code for the 4map6 Tunnel in "BGP Tunnel Encapsulation Attribute Tunnel Types" registry.

3) Two codes for the "Address Origin Type" and "Forwarding Type" Sub-TLVs in "BGP Tunnel Encapsulation Attribute Sub-TLVs" registry.

All the codes above use this document as the reference.

9. Acknowledgement

The authors would like to thank Jeffrey Haas, Susan Hares, Robert Raszuk, Changwang Lin, Yingzhen Qu, Nan Geng, Gyan Mishra, Jingrong Xie, Acee Lindem, Tom Petch, Richard Huang, Cheng Li, Jie Dong, Eduard Metz, Shunwan Zhang, Linjian Song, Weiqiang Cheng, Paolo Volpato, Sheng Jiang, Giuseppe Fioccola, Shuping Peng, Di Ma, Ran Pang, Tianran Zhou, Linda Dunbar for their review and comments.

10. References

10.1. Normative References

[I-D.ietf-v6ops-framework-md-ipv6only-underlay] Xie, C., Ma, C., Li, X., Mishra, G. S., and T. Graf, "Framework for Multi-domain IPv6-only Underlay Network and IPv4-as-a-Service", Work in Progress, Internet-Draft, draft-ietf-v6ops-framework-md-ipv6only-underlay-15, 27 November 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-v6ops-framework-md-ipv6only-underlay-15>.

[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>.

[RFC6368] Marques, P., Raszuk, R., Patel, K., Kumaki, K., and T. Yamagata, "Internal BGP as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 6368, DOI 10.17487/RFC6368, September 2011, <https://www.rfc-editor.org/info/rfc6368>.

[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, <https://www.rfc-editor.org/info/rfc7606>.

[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>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

10.2. Informative References

[IANA-BGP-TUNNEL-ENCAP] IANA, "Border Gateway Protocol (BGP) Tunnel Encapsulation", <https://www.iana.org/assignments/bgp-tunnel-encapsulation/>.

Appendix A. Contributors

The following people have contributed to this document:

Congxiao Bao
CERNET Center/Tsinghua University

Email: congxiao@cernet.edu.cn

Linjian Song
Alibaba Cloud
Email: linjian.slj@alibaba-inc.com

Chenhao Ma
China Telecom
Email: Machh@chinatelecom.cn

Appendix B. IPv6-only DCN for AI-infra Fabric

There is enormous "East-West" traffic inside the data center network, which are the flows between DC devices and applications. Upgrading the DCN network firstly to dual-stack, then IPv6-only is nontrivial. One example is building AI-infra fabric on IPv6 only fabric which reduce data plane encapsulation overhead, simplify forwarding chip’s feature and improve data plane performance.

When DCN plans to transits from dual stack to IPv6-only, it is impossible to be done overnight. Considerations and plans should be made supporting legacy IPv4 servers and applications when the DCN is IPv6-only. The IPv6-only framework proposed in [I-D.ietf-v6ops-framework-md-ipv6only-underlay] provides availability for IPv4 service when the underlay Networks upgraded to IPv6-only.

As shown in Figure 6, Host 1 and Host 2 are legacy servers with only IPv4 capability. Traffic between Host 1 and Host 2 are carried by IPv6 network in the DCN. The access switch(ASW) have the function of ADPT which learns IPv4/IPv6 mapping rules and delivers the IPv4 service in IPv6-only network.

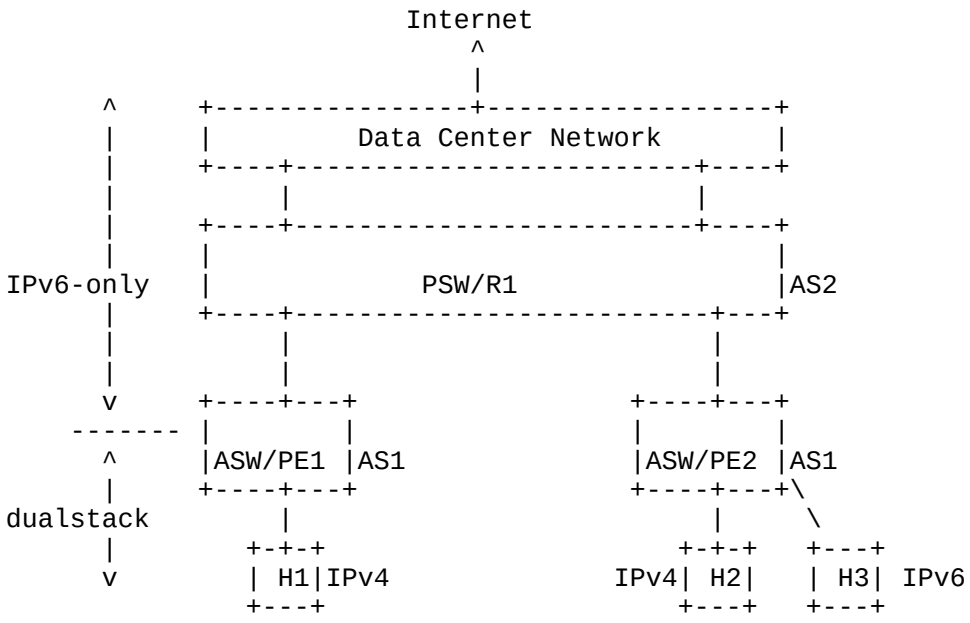

Figure 6: IPv6-only DCN for AI infra fabric

Authors' Addresses

Chongfeng Xie
China Telecom
Beiqijia Town, Changping District
Beijing
102209
China
Email: xiechf@chinatelecom.cn

Guozhen Dong
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

Guoliang Han
Indirection Network Inc.
Email: guoliang.han@indirectionnet.com

Zhongfeng Guo
Alibaba Cloud
Wangjing Qiyang Rd, Chaoyang District
Beijing
100102
China
Email: guozhongfeng.gzf@alibaba-inc.com