

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 30 July 2022

E. Horley
T. Coffeen
S. Hogg
HexaBuild
N. Buraglio
C. Cummings
Energy Sciences Network
K. Myers
IP ArchiTechs
R. White
Juniper Networks
26 January 2022

Expanding the IPv6 Lab Use Space
draft-horley-v6ops-lab-02

Abstract

To reduce the likelihood of addressing conflicts and confusion between lab deployments and non-lab (i.e., production) deployments, an IPv6 unicast address prefix is reserved for use in lab, proof-of-concept, and validation networks as well as for any similar use case. This document describes the use of the IPv6 address prefix 0200::/7 as a prefix reserved for this purpose (repurposing the deprecated OSI NSAP-mapped prefix).

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 30 July 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
2. New Lab IPv6 Address Prefix	3
3. Operational Implications	3
4. IANA Considerations	4
5. Security Considerations	4
6. Acknowledgements	4
7. References	4
7.1. Normative References	4
7.2. Informative References	5
Authors' Addresses	5

1. Introduction

The address architecture for IPv6 ([RFC4291]) does not explicitly define any prefixes allocated exclusively for lab use, nor is such address space allocated in [RFC6890] or in [RFC8200]. While lab deployments could potentially use IPv6 address prefixes typically assigned and configured in non-lab network, the use of such addressing in lab environments may create addressing conflicts and operational confusion. For instance, designing labs utilizing ULA `fc00::/7` [RFC4193] is problematic due to the random global ID requirement preventing hierarchical network prefix design possibilities. Further, default address selection behavior [RFC6724] by end nodes may result in a depreferencing of such addresses and prevent lab deployments from accurately modeling their desired non-lab equivalents.

To resolve these problems involved in building large-scale lab networks, and pre-staging, or automating large-scale networks for deployment, this document allocates the IPv6 address prefix `0200::/7` for these purposes.

The goal is to allow organization to share working lab configuration files (with little or no need of modification) to be deployed in a third party lab environment like,

public and private clouds,

virtualization or hosting environments,

and in other networks like Service Providers, Enterprise, Government, IoT, and Energy,

all with the knowledge that the lab GUA address space will perform the same as any GUA but with the added knowledge that filtering will be used to protect accidental leaks to the Internet.

The following criteria is for selecting the lab prefix:

The precedence for the lab prefix should no be lower than the GUA prefix as defined in [RFC6724] (unlike ULA). Reduce the operational impacts to IANA and the RIR's in selecting lab prefix space.

2. New Lab IPv6 Address Prefix

The prefix reserved for lab and testing purposes is 0200::/7.

3. Operational Implications

This space SHOULD NOT be employed for addressing use cases which are already defined in other RFCs, such as addresses set apart for documentation, testing, etc.

Enterprise and large-scale networks have some specific criteria around building and validating prior to deployment. The issues with ULA for infrastructure modeling and labbing at the host level are more impactful in large enterprises. This is due to the increased focus on large-scale hosts, servers, and apps testing. Also, it is likely that both GUA and ULA may co-exist (or are planned) and reconfiguring lab hosts and networks isn't practical or desirable due to inconsistent results for host preference due to [RFC6724] behavior.

Most large enterprises strive to build lab, dev, and QA environments that reflect production as accurately as possible. This is a fairly straightforward way to avoid disparity between production and non-production. Enterprise environments are an area that need increased IPv6 adoption. In an effort to make it easier to model a global enterprise and to avoid the pitfalls of ULA de-preferenced host behavior or squatting on other IPv6 space, a specific IPv6 lab prefix is being assigned.

Because this address prefix has previously been used for the OSI NSAP-mapped prefix set in [RFC4048] and [RFC4548], and deprecated, this address prefix is already limited in its usability. In addition, the address prefix was returned to IANA and is available to be marked for lab or other purposes.

This assignment implies that IPv6 network operators SHOULD add this address prefix to the list of non-routable IPv6 address space, and if packet filters are deployed, then this address prefix SHOULD be added to packet filters. This is not a local-use address prefix so these filters may be used in both local and public contexts.

4. IANA Considerations

IANA is to record the reservation of the IPv6 global unicast address prefix 0200::/7 as a lab-only prefix in the IPv6 address registry. No end party is to be assigned this address.

5. Security Considerations

The addresses assigned for lab and staging use SHOULD be filtered as noted above.

Setting aside address space for lab and staging use, and adding this address space to common filters to prevent destinations in this space from being routed in production networks (including the global Internet) improves security by preventing the leakage of prefixes used for testing into production environments. As such, setting aside this space improves the overall security posture of the Internet.

6. Acknowledgements

The authors acknowledge the work of Bob Hinden and Stephen Deering, in authoring the protocol standard and the addressing architecture for IPv6.

7. References

7.1. 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>>.

- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

7.2. Informative References

- [RFC3515] Sparks, R., "The Session Initiation Protocol (SIP) Refer Method", RFC 3515, DOI 10.17487/RFC3515, April 2003, <<https://www.rfc-editor.org/info/rfc3515>>.
- [RFC4048] Carpenter, B., "RFC 1888 Is Obsolete", RFC 4048, DOI 10.17487/RFC4048, April 2005, <<https://www.rfc-editor.org/info/rfc4048>>.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, DOI 10.17487/RFC4193, October 2005, <<https://www.rfc-editor.org/info/rfc4193>>.
- [RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <<https://www.rfc-editor.org/info/rfc4291>>.
- [RFC4548] Gray, E., Rutenmiller, J., and G. Swallow, "Internet Code Point (ICP) Assignments for NSAP Addresses", RFC 4548, DOI 10.17487/RFC4548, May 2006, <<https://www.rfc-editor.org/info/rfc4548>>.
- [RFC5180] Popoviciu, C., Hamza, A., Van de Velde, G., and D. Dugatkin, "IPv6 Benchmarking Methodology for Network Interconnect Devices", RFC 5180, DOI 10.17487/RFC5180, May 2008, <<https://www.rfc-editor.org/info/rfc5180>>.
- [RFC6724] Thaler, D., Ed., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, DOI 10.17487/RFC6724, September 2012, <<https://www.rfc-editor.org/info/rfc6724>>.
- [RFC6890] Cotton, M., Vegoda, L., Bonica, R., Ed., and B. Haberman, "Special-Purpose IP Address Registries", BCP 153, RFC 6890, DOI 10.17487/RFC6890, April 2013, <<https://www.rfc-editor.org/info/rfc6890>>.

Authors' Addresses

Ed Horley
HexaBuild

Email: ed@hexabuild.io

Tom Coffeen
HexaBuild

Email: tom@hexabuild.io

Scott Hogg
HexaBuild

Email: scott@hexabuild.io

Nick Buraglio
Energy Sciences Network

Email: buraglio@es.net

Chris Cummings
Energy Sciences Network

Email: chriscummings@es.net

Kevin Myers
IP ArchiTechs

Email: kevin.myers@iparchitech.com

Russ White
Juniper Networks

Email: russ@riw.us

V6OPS
Internet-Draft
Intended status: Informational
Expires: September 8, 2022

G. Fioccola
P. Volpato
Huawei Technologies
N. Elkins
Inside Products
J. Palet Martinez
The IPv6 Company
G. Mishra
Verizon Inc.
C. Xie
China Telecom
March 7, 2022

IPv6 Deployment Status
draft-ietf-v6ops-ipv6-deployment-05

Abstract

This document provides an overview of IPv6 deployment status and a view on how the transition to IPv6 is progressing among network operators and enterprises. It also aims to analyze the related challenges and therefore encourage actions and more investigations in those areas where the industry has not taken a clear and unified approach.

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 September 8, 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. IPv4 vs IPv6: The Global Picture	5
2.1. IPv4 Address Exhaustion	5
2.2. IPv6 Users	6
2.3. IPv6 Web Content	7
2.4. IPv4 addresses per capita and IPv6 status	7
3. A Survey on IPv6 Deployments	10
3.1. IPv6 Allocations and Networks	10
3.2. IPv6 among Network Operators	12
3.3. IPv6 among Enterprises	14
3.3.1. Government and Universities	15
3.4. Observations on Industrial Internet	17
3.5. Observations on Content and Cloud Service Providers	17
3.6. Application Transition	18
4. Towards an IPv6 Overlay Service Design	18
4.1. IPv6 introduction	19
4.2. IPv6-only Service Delivery	20
5. IPv6-only Underlay Network Deployment	21
6. IPv6 Benefits	23
7. Common IPv6 Challenges	25
7.1. Transition Choices	25
7.1.1. Service Providers	25
7.1.2. Enterprises and Industrial Internet	26
7.1.3. Cloud and Data Centers	27
7.1.4. CEs and user devices	28
7.2. Government and Regulators	28
7.3. Network Management and Operations	29
7.4. Performance	29
7.4.1. IPv6 packet loss and latency	30
7.4.2. Customer Experience	30
7.5. IPv6 security	31
7.5.1. Protocols security issues	32
7.5.2. IPv6 Extension Headers and Fragmentation	33
8. Security Considerations	33
9. Contributors	34
10. Acknowledgements	34

11. IANA Considerations	34
12. References	34
12.1. Normative References	34
12.2. Informative References	36
Appendix A. Summary of Questionnaire and Replies for network operators	42
Appendix B. Summary of Questionnaire and Replies for enterprises Authors' Addresses	44 45

1. Introduction

[RFC6036] described IPv6 deployment scenarios adopted or foreseen by a number of Internet Service Providers (ISPs) who responded to a technical questionnaire in early 2010. In doing that, [RFC6036] provided practices and plans expected to take place in the following years. Since the publication of [RFC6036], several other documents contributed to discuss the transition to IPv6 in operational environments. To name a few:

- [RFC6180] discussed IPv6 deployment models and transition mechanisms, recommending those that showed to be effective in operational networks.
- [RFC6883] provided guidance and suggestions for Internet content providers and Application Service Providers (ASPs).
- [RFC7381] introduced the guidelines of IPv6 deployment for enterprises.

It is worth mentioning here also [RFC6540] that not only recommended the support of IPv6 to all IP-capable nodes, but it was referenced in the IAB Statement on IPv6 [IAB], which represented a major step in driving the IETF as well as other Standard Developing Organizations (SDOs) to assume the use of IPv6 in their works.

In more recent times, organizations such as ETSI provided more contributions to the use of IPv6 in operational environments, targeting IPv6 in different industry segments. As a result, [ETSI-IP6-WhitePaper], was published to provide an updated view on the IPv6 best practices adopted so far, in particular in the ISPs domain.

Considering all of the above, and after more than ten years since the publication of [RFC6036] it may be interesting to verify where the transition of the Internet to IPv6 currently stands, what major steps have been accomplished and what is still missing. Some reasons justify such questions:

- In some areas, the lack of publicly available IPv4 addresses forced both carriers and content providers to shift to IPv6 to support the introduction of new applications, in particular in wireless networks.
- Some governmental actions took place to encourage or even enforce, at different degrees, the adoption of IPv6 in certain countries.
- Looking at the global adoption of IPv6, this seems to have reached a threshold that justifies speaking of native, end-to-end IPv6 connectivity (between a user's device and a content on a site) at the IPv6 service layer.

This document intends to explode such statements, providing a survey of the status of IPv6 deployment and highlighting both the achievements and remaining obstacles in the transition to IPv6-only networks. The target is to give an updated view of the practices and plans already described in [RFC6036], to encourage further actions and more investigations in those areas that are still under discussion, and to present the main incentives for the adoption of IPv6. The expectation is that this process may help to understand what is missing and how to improve the current IPv6 deployment strategies of network operators, enterprises, content and cloud service providers.

The initial section of this document reports some data about the status of IPv6. The exhaustion of IPv4 as well as the measured adoption of IPv6 at the users' and the content's side will be discussed. Comparing both IPv4 and IPv6, this latter has a higher growth rate. While this fact alone does not permit to conclude that the definitive transition to IPv6 is undergoing, at least testifies that a portion of the ICT industry has decided to invest and deploy IPv6 at large.

The next section provides a survey of IPv6 deployments in different environments, including ISPs, enterprises, cloud providers and universities. Data from some well-known analytics will be discussed. In addition, two independent polls among network operators and enterprises will also be presented.

Then, a section on IPv6 overlay service design will describe the IPv6 transition approaches for Mobile BroadBand (MBB), Fixed BroadBand (FBB) and Enterprise services. At present, Dual-Stack (DS) is the most deployed solution for IPv6 introduction, while 464XLAT and Dual-Stack Lite (DS-Lite) seem the preferred ones for those players that have already enabled IPv6-only service delivery. A section on IPv6

underlay network deployment will also focus on the common approaches for the transport network.

The last parts of the document will analyze the incentives brought by IPv6 as well as the general challenges to be faced to move forward in the transition. Specific attention will be given to operational, performance and security issues. All these considerations will be input for the final section of the document that aims to highlight the areas still requiring improvement and some actions that the industry might consider to favor the adoption of IPv6.

2. IPv4 vs IPv6: The Global Picture

This section deals with some key questions related to IPv6, namely the status of IPv4 exhaustion, often considered as one of the triggers to switch to IPv6, the number of IPv6 end users, a primary measure to sense IPv6 adoption, and the percentage of websites reachable over IPv6. The former is constantly measured by the Regional Internet Registries (RIRs) and the next subsection provides an indication of where we currently stand. The utilization of IPv6 at both the end user's side and the content's side has also been monitored by several institutions worldwide as these two parameters provide a first-order indication on the real adoption of IPv6.

2.1. IPv4 Address Exhaustion

According to [CAIR] there will be 29.3 billion networked devices by 2023, up from 18.4 billion in 2018. This poses the question on whether the IPv4 address space can sustain such a number of allocations and, consequently, if this is affecting the process of its exhaustion. The answer is not straightforward as many aspects have to be considered.

On one hand, the RIRs are reporting scarcity of available and still reserved addresses. Table 3 of [POTAR001] shows that the available pool of the five RIRs counts 5.2 million IPv4 addresses, while the reserved pool includes another 12 million, for a total of "usable" addresses equal to 17.3 million (-5.5% year over year, comparing 2021 against 2020). The same reference, in table 1, shows that the total IPv4 allocated pool equals to 3.685 billion addresses (+0.027% year over year). The ratio between the "usable" addresses and the total allocated brings to 0.469% of remaining space (from 0.474% at the end of 2020).

On the other, [POTAR001] again highlights the role of both NAT and the address transfer to counter the IPv4 exhaustion. NAT systems well fit in the current client/server model used by most of the available Internet applications, with this phenomenon amplified by

the general shift to cloud. Anyway, it should be noted that, in some cases, private address space cannot provide adequate address and the reuse of addresses may make the network even more complex. The transfer of IPv4 addresses also contributes to mitigate the need of addresses. As an example, [IGP-GT] and [NRO] show the amount of transfers to recipient organizations in the different regions. Cloud Service Providers (CSPs) appear to be the most active in buying available addresses to satisfy their need of providing IPv4 connectivity to their tenants. But, since each address blocks of Internet is licensed by a specific resource-holder and stored for the verification of the authenticity, frequent address transfer may affect the global assignment process.

2.2. IPv6 Users

The count of the IPv6 users is the key parameter to get an immediate understanding of the adoption of IPv6. Some organizations constantly track the usage of IPv6 by aggregating data from several sources. As an example, the Internet Society constantly monitors the volume of IPv6 traffic for the networks that joined the WorldIPv6Launch initiative [WIPv6L]. The measurement aggregates statistics from organizations such as [Akm-stats] that provides data down to the single network level measuring the number of hits to their content delivery platform. For the scope of this document, we follow the approach used by APNIC to quantify the adoption of IPv6 by means of a script that runs on a user's device [CAIDA]. To give a rough estimation of the relative growth of IPv6, the next table aggregates the total number of estimated IPv6-capable users at January 2022, and compares it against the total Internet users, as measured by [POTAROO2].

	Jan 2018	Jan 2019	Jan 2020	Jan 2021	Jan 2022	CAGR
IPv6	513.07	574.02	989.25	1,136.28	1,207.61	23.9%
World	3,410.27	3,470.36	4,065.00	4,091.62	4,093.69	4.7%
Ratio	15.0%	16.5%	24.3%	27.8%	29.5%	18.4%

Figure 1: IPv6-capable users against total (in millions)

Two figures appear: first, the IPv6 Internet population is growing with a two-digits Compound Annual Growth Rate (CAGR), and second, the ratio IPv6 over total is also growing steadily.

2.3. IPv6 Web Content

[W3Tech] keeps track of the use of several technical components of websites. The utilization of IPv6 for websites is shown in the next table.

Worldwide Websites	Jan 2018	Jan 2019	Jan 2020	Jan 2021	Jan 2022	CAGR
% of IPv6	11.4%	13.3%	15.0%	17.5%	20.6%	15.9%

Figure 2: Usage of IPv6 in websites

Looking at the growth rate, it may appear not particularly high. It has to be noted, though, that not all websites are equal. The largest content providers, which already support IPv6, generate a lot more IPv6-based content than small websites. [Csc6lab] measured at the beginning of January 2022 that out of the world top 500 sites ranked by [Alx], 203 are IPv6-enabled (+3.6% from January 2021 to January 2022). If we consider that the big content providers (such as Google, Facebook, Netflix) generate more than 50% of the total mobile traffic [SNDVN], and in some cases even more up to 65% ([ISOC1] [HxBld]), the percentage of content accessible over IPv6 is clearly more relevant than the number of enabled IPv6 websites.

Related to that, a question that sometimes arises is whether the content stored by content providers would be all accessible on IPv6 in the hypothetical case of a sudden IPv4 switch-off. Even if this is pure speculation, the numbers above may bring to state that this is likely the case. This would reinforce the common thought that, in quantitative terms, most of content is accessible via IPv6.

2.4. IPv4 addresses per capita and IPv6 status

The IPv4 addresses per capita ratio measures the quantity of IPv4 addresses allocated to a given country divided by the population of that country. It is a theoretical ratio, anyhow it provides an indication of the imbalanced distribution of the IPv4 addresses worldwide. It clearly derives from the allocation of addresses made in the early days of the Internet by the most advanced countries.

The sources for measuring the IPv4 addresses per capita ratio are the allocations done by the RIRs and the statistics about the world population. In our case, we take the distribution files of [POTAROO2], which summarize the most useful information. We then

compare the obtained ratio against the measured adoption of IPv6. As explained in section 2.2, this is expressed as the number of IPv6 capable users over the total Internet population of the considered country. The result is shown in the following table, where some of countries with the highest number of IPv4 allocations are reported. The table is ordered based on the IPv4 addresses per capita ratio.

Country	IPv4 per capita	IPv6 deployment
United States of America	4.89	47.1%
Sweden	2.97	11.8%
Netherlands	2.93	35.5%
Switzerland	2.75	34.9%
Republic of Korea	2.19	17.1%
Australia	1.91	28.8%
Canada	1.85	32.4%
United Kingdom	1.65	33.2%
Belgium	1.62	62.0%
Japan	1.50	36.7%
Germany	1.48	53.0%
France	1.27	42.1%
Austria	1.24	29.2%
Italy	0.91	4.7%
Spain	0.69	3.0%
Poland	0.55	14.7%
Brazil	0.41	38.7%
Russian Federation	0.31	9.7%
China (*)	0.24	60.1%
India	0.03	76.9%

Figure 3: IPv4 per capita and IPv6 deployment

(*) The IPv6 deployment information in China is derived from [CN-IPv6].

It is clear that a direct correlation between the two measures is not always straightforward. One may expect that a low IPv4 addresses per capita ratio should correspond to high IPv6 deployment. This is true for India and, relatively, for Brazil, but other countries such as the Russian Federation, Poland, Spain and Italy are still quite dependent on IPv4.

The opposite effect should apply for countries with a high value for same ratio. While this is true for Sweden and the Republic of Korea, which have a relatively low IPv6 deployment, for other countries at the top of the list this does not apply. The USA, Germany, France, Belgium all have a good level of IPv6 deployment, while other countries such the Netherlands, Switzerland, the UK, Canada, Australia, Japan are just following with around one third of their Internet population using IPv6.

The reasons for having high or medium-to-high IPv6 deployment may be quite different from country to country. For example, in West European countries such as Belgium, France, Germany the outcome depends on a mix of government and regulation activities which incentivized the adoption of IPv6 in the recent years. In some cases it has been industry self-regulation which created the traction for a bigger IPv6 adoption.

The IPv6 adoption in USA, China and India comes from different needs. Mobile operators have anticipated the usage of IPv6 to cope with the lack of available addresses (India) and to support new services and applications. In addition to that, all National governments have issued specific policies to bring IPv6 deployment forward.

3. A Survey on IPv6 Deployments

Right after the count of the IPv6 users, it is fundamental to understand the status of IPv6 in terms of concrete adoption in operational networks. This section deals with the status of IPv6 among carriers, service and content providers, enterprises and research institutions.

3.1. IPv6 Allocations and Networks

RIRs are responsible for allocating IPv6 address blocks to ISPs, LIRs (Local Internet Registries) as well as enterprises or other organizations. An ISP/LIR will use the allocated block to assign addresses to their end users. For example, a mobile carrier will assign one or several /64 prefixes to the User Equipment (UE).

Several analytics are available from the RIRs. Here we are interested to those relevant to IPv6. The next table shows the amount of individual allocations, per RIR, in the time period 2017-2021 [APNIC2].

Registry	Dec 2017	Dec 2018	Dec 2019	Dec 2020	Dec 2021	Cumulated	CAGR
AFRINIC	112	110	115	109	136	582	51%
APNIC	1,369	1,474	1,484	1,498	1,392	7,217	52%
ARIN	684	659	605	644	671	3,263	48%
LACNIC	1,549	1,448	1,614	1,801	730	7,142	47%
RIPE NCC	2,051	2,620	3,104	1,403	2,542	11,720	55%
Total	5,765	6,311	6,922	5,455	5,471	29,924	51%

Figure 4: IPv6 allocations worldwide

Overall, the trend is positive, witnessing the vivacity around IPv6. The decline of IPv6 allocations in 2020 (remarkable for RIPE NCC) and 2021 (in LACNIC), could be explained with the COVID-19 measures that may have affected the whole industry.

[APNIC2] also compares the number of allocations for both address families. Differently from what happened in 2020, when CAGR was higher for the IPv6 allocations, in 2021 IPv4 stayed a little ahead (53.6% versus 50.9%).

Address family	Dec 2017	Dec 2018	Dec 2019	Dec 2020	Dec 2021	Cumulated	CAGR
IPv6	5,765	6,311	6,922	5,455	5,471	29,924	50.9%
IPv4	8,091	9,707	13,112	6,263	7,829	45,002	53.6%

Figure 5: Allocations per address family

As noted, the pandemic may have influenced the ICT industry including the dynamic of the address allocations. The number of IPv4 allocations in 2021 includes many allocations of small address ranges (e.g. /24) [APNIC2]. On the contrary, a single IPv6 allocation is

large enough to cope with the need of an operators for many years. After an operator receives an IPv6 /30 or /32 allocation it is unlikely that a new request of addresses is repeated in the short term. Hence the two CAGR values in the table should not be compared directly as the weight of the allocations is different.

The next table is based on [APNIC3], [APNIC4] and shows the percentage of Autonomous System (AS) numbers supporting IPv6 compared to the total ASes worldwide. The number of IPv6-capable ASes increased from 24.3% in January 2018 to 38.7% in January 2022. This equals to 18% CAGR for IPv6-enabled networks, highlighting how IPv6 is growing faster than IPv4, since the CAGR value for the total of IPv6 and IPv4 networks grew of just 5%.

Advertised ASN	Jan 2018	Jan 2019	Jan 2020	Jan 2021	Jan 2022	CAGR
IPv6-capable	14,500	16,470	18,650	21,400	28,140	18%
Total ASN	59,700	63,100	66,800	70,400	72,800	5%
Ratio	24.3%	26.1%	27.9%	30.4%	38.7%	

Figure 6: Percentage of IPv6-capable ASes

The tables above provide an aggregated view of the allocations dynamic. Apart from the recent times influenced by the pandemic, the general trend related to IPv6 adoption is positive. What the aggregated view does not tell us is the split between the different types of organizations. The next sections of this chapter will zoom into each specific area to highlight the relative status.

3.2. IPv6 among Network Operators

Only a few public references describing the status of IPv6 in specific networks are available. An example is the case of Reliance Jio [RlncJ]. To understand the degree of adoption of IPv6 in the operators' domain, it is necessary to consult the data provided by those organizations that constantly track the usage of IPv6. Among the others, we have the Internet Society that constantly monitors the volume of IPv6 traffic for the networks that joined the WorldIPv6Launch initiative [WIPv6L] and Akamai [Akm-stats] that collects statistics both at a country level and at the single operator's network measuring the number of hits to their content delivery platform. In addition to them, the RIRs also provide

detailed information about the prefixes allocated and the ASes associated to each operator. Overall, the vast majority of the operators worldwide have enabled IPv6 and provide IPv6-based services even if the degree of adoption varies quite greatly based on local market demand, regulatory actions, and political decisions (e.g. [RIPE3] to look at the relative differences across the European market).

As it was proposed at the time of [RFC6036], also in the case of this document a survey was submitted to a group of service providers in Europe (see Appendix A for the complete poll), to understand the details about their plans about IPv6 and their technical preferences towards its adoption. Such poll does not pretend to give an exhaustive view on the IPv6 status, but to integrate the available data with some insights that may be relevant to the discussion.

The poll reveals that the majority of the operators interviewed has plans concerning IPv6 (79%). Of them, 60% has already ongoing activities, while 33% is expected to start activities in a 12-months time-frame. The transition to IPv6 involves all business segments: mobile (63%), fixed (63%), and enterprises (50%).

The reasons to move to IPv6 vary. The majority of the operators that do have a plan for IPv6 perceives issues related to IPv4 depletion and prefer to avoid the use of private addressing schemes (48%) to save the NAT costs. Global IPv4 address depletion and the run out of private address space recommended in [RFC1918] are reported as the important drivers for IPv6 deployment. In some cases, the adoption of IPv6 is driven by innovation strategy (as the enabler of new services, 13%) or is introduced because of 5G/IoT, which play the role of business incentive to IPv6 (20%). In a few cases, respondents highlight the availability of National Regulatory policies requiring to enable IPv6 together with the launch of 5G (13%). Enterprise customers demand is also a reason to introduce IPv6 (13%).

From a technical preference standpoint, Dual-Stack is the most adopted solution, both in wireline (59%) and in cellular networks (39%). In wireline, the second most adopted mechanism is DS-Lite (19%), while in cellular networks the second preference goes to 4G4XLAT (21%).

In the majority of the cases, the interviewed operators do not see any need to transition their network as a whole. They consider to touch or to replace only what it is needed. CE (47%), BNG (20%), CGN devices (33%), mobile core (27%) are the components that may be affected by transition or replacement. It is interesting to see that most of the network operators have no major plans to transition the

transport network (metro and backbone) soon, since they do not see business reasons. It seems that there is no pressure to move to native IPv6-only forwarding in the short term, anyway the future benefit of IPv6 may justify the shift in the long term.

More details about the answers received can be found in Appendix A.

3.3. IPv6 among Enterprises

As described in [RFC7381], enterprises face different challenges than operators. Some publicly available statistics also show that the deployment of IPv6 lags behind other sectors.

[NST_1] provides estimations on deployment status of IPv6 for more than 1000 second level domains such as example.com, example.net or example.org belonging to organizations in the United States. The measurement encompasses many industries, including telecommunications, so that the term "enterprises" is a bit loose to this extent. In any case, it provides a first indication of IPv6 adoption in several US industry sectors. The analysis tries to infer whether IPv6 is supported by looking from "outside" a company's network. It takes into consideration the support of IPv6 to external services such as Domain Name System (DNS), mail and website. Overall, for around the 66.85% of the considered domains there is an active DNS Name Server (NS) v6 record but only 6.3% have IPv6 support for their websites and 21.2% have IPv6-based mail services as of January 2022.

[BGR_1] have similar data for China. The measurement considers 478 second or third level domains such as example.cn or example.com.cn. 74% have IPv6 support for DNS, 0% have IPv6-based mail services, 20% have IPv6-based websites.

[CNLABS_1] provides the status in India. Of the 104 measured domains, 51.9% are IPv6-enabled, 15.4% have IPv6 support in DNS, 16.3% have IPv6 support for the mail service.

A poll submitted to a group of large enterprises in North America (see Appendix B) show that the operational issues are likely to be more critical than for operators.

Looking at current implementations, almost one third has dual-stacked networks, while 20% declares that portions of their networks are IPv6-only. 35% of the enterprises are stuck at the training phase. In no cases the network is fully IPv6-based.

Speaking of training, the most critical needs are in the field of IPv6 security and IPv6 troubleshooting (both highlighted by the two thirds of respondents), followed by IPv6 fundamentals (57.41%).

Coming to implementation, the three areas of concern are IPv6 security (31.48%), training (27.78%), application conversion (25.93%). Interestingly, 33.33% of respondents think that all three areas are all simultaneously of concern.

The full poll is reported in Appendix B.

3.3.1. Government and Universities

This section focuses specifically on governments and academia, due to the relevance of both domains in the process of IPv6 adoption. The already mentioned organizations that estimates the IPv6 status provide a deep focus on IPv6 in the network domains associated with governmental and education-related agencies.

As far as governmental institutions or agencies are concerned, [NST_2] provides analytics on the degree of support that IPv6 gives to DSN, mail and websites across 1283 second level domains associated with US Federal agencies. These domains are in the form of example.gov or example.fed. The script used by [NST_2] have also been employed to measure the same analytics in other countries. For China [BGR_2] looks at 52 third level domains such as example.gov.cn. [CNLABS_2] provides statistics for 618 Indian government-related domains (ranging from the second to the fifth level). [IPv6Forum] analyzes 19 governmental domains connected to either the European Union or its member States.

Country	Domains analyzed	DNS	Mail	Website
China	52	0.0%	0.0%	98.1%
European Union (*)	19	47.4%	0.0%	21.1%
India	618	7.6%	6.5%	7.1%
United States of America	1283	87.1%	14.0%	51.7%

Figure 7: IPv6 support for external-facing services across governmental institutions

(*) Both EU and European governments domains are considered.

Looking at the USA, the support given by IPv6 to the services is higher than that in the enterprise sector discussed in the previous section. This is likely to depend on the directions set by [US-CIO] through the mandate to transition to IPv6. In the case of India, the degree of support seems still quite low. This is also true for China, with the notable exception of the percentage of IPv6-enabled government-related organizations websites.

Similar statistics are also available for higher education. [NST_3] measures the data coming from 346 second level domains of universities in the US, such as example.edu. [BGR_3] looks at 111 Chinese education-related domains such as example.edu.cn. [CNLABS_3] analyzes 100 domains in India (mostly third level), while [IPv6Forum] lists 118 universities in the European Union (second level).

Country	Domains analyzed	DNS	Mail	Website
China	111	36.9%	0.0%	77.5%
European Union	118	83.9%	43.2%	35.6%
India	100	31.0%	54.0%	5.0%
United States of America	346	49.1%	19.4%	21.7%

Figure 8: IPv6 support for external-facing services across universities

Overall, the universities have wider support of IPv6-based services if compared with the other sectors. Apart from a couple of exceptions (e.g. the support of IPv6 mail in China and IPv6 web sites in India), the numbers shown in the table above indicate a good support of IPv6.

3.4. Observations on Industrial Internet

There are potential advantages for implementing IPv6 for IIoT (Industrial Internet of Things) applications, in particular the large IPv6 address space, the automatic IPv6 configuration and resource discovery.

However, there are still many obstacles that prevent its pervasive use. The key problems identified are the incomplete or immature tool support, the dependency on manual configuration and the poor knowledge of the IPv6 protocols among insiders. To advance and ease the use of IPv6 for smart manufacturing systems and IIoT applications in general, a generic approach to remove these pain points is therefore, highly desirable.

3.5. Observations on Content and Cloud Service Providers

Both the number of addresses required to connect all of the virtual and physical elements in a Data Center and the necessity to overcome the limitation posed by [RFC1918] have been the drivers to adopt IPv6 in several CSP networks.

Several public references, as reported in Section 7.1.3, discuss how most of the major players find themselves at different stages in the transition to IPv6-only in their Data Center (DC) infrastructure. In some cases, the transition already happened and the DC infrastructure of these hyperscalers is completely based on IPv6. This can be considered a good sign because the end-to-end connectivity between a client (e.g. an application on a smartphone) and a server (a Virtual Machine in a DC) may be based on IPv6.

3.6. Application Transition

The preliminary step to take full benefit of the IPv6 capabilities is to write or adapt the application software for use in IPv6 networks (see, as an example, [ARIN-SW]).

It is worth mentioning Happy Eyeballs [RFC6555] and Happy Eyeballs 2 [RFC8305] as a major aspect of application transition and porting to IPv6. All host and network router OS's by default prefer IPv6 over IPv4.

Happy Eyeballs plays an important role and is extremely helpful when applications are migrated from IPv4 to IPv6 having a single DNS name with both A and AAAA record. In this way all applications can migrate to IPv6 as IPv6 is preferred over IPv4 and so all IPv6 dual-stacked hosts can communicate using same URL to the server via IPv6 and all IPv4 hosts can as well use the same URL to communicate via IPv4. Eventually when all host endpoints are dual-stacked, the application servers can migrate from dual stack to IPv6-only. For external connectivity to the internet, web proxy can be used providing 6to6 or 4to6 proxy to the internet. This allows application servers on the internal network to change to IPv6-only once all host endpoints are all dual-stacked.

At the current stage, the full support of IPv6 is not yet complete [Wikipedia], as issues remains in particular for applications known not to work properly behind NAT64.

4. Towards an IPv6 Overlay Service Design

This section reports the most deployed approaches for the IPv6 transition in MBB, FBB and enterprise.

Two approaches are usually considered [ETSI-IP6-WhitePaper], namely: (1) IPv6 introduction, and (2) IPv6-only. Depending on their specific plans and requirements, operators may consider either of the two or both. The usual approach sees them as two sequential stages, i.e. deal with IPv6 introduction first and then move to IPv6-only. In some cases, operators may instead jump directly to IPv6-only to

avoid the operational burden of a transient phase. IPv6 introduction aims at delivering the service in a controlled manner, where the traffic volume of IPv6-based services is minimal. When the service conditions change, e.g. when the traffic grows beyond a certain threshold, then the move to IPv6-only may occur. In this latter case, the service is delivered solely on IPv6, including the traffic originated from IPv4-based nodes. For this reason, the IPv6-only stage is also called IPv4aaS (IPv4 as a Service).

The consolidated approach foresees to enable IPv6 in the network (sometimes referred to as the underlay) and move progressively to the service layer. Recently, the attention has shifted to enabling IPv6 at the service layer (the overlay) leaving the transition of the network to IPv6 at a later stage. This relates to the increased adoption of the transition mechanisms described in this section.

4.1. IPv6 introduction

In order to enable the deployment of an IPv6 service over an underlay IPv4 architecture, there are two possible approaches:

- o Enabling Dual-Stack [RFC4213] at the Customer Edge router (CE)
- o IPv6-in-IPv4 tunneling, e.g. with IPv6 Rapid Deployment (6rd) or Generic Routing Encapsulation (GRE).

Based on information provided by operators with the answers to the poll (Appendix A), Dual-Stack appears to be currently the most widely deployed IPv6 solution, for MBB, FBB and enterprises, accounting for about 50% of all IPv6 deployments (see both Appendix A and the statistics reported in [ETSI-IP6-WhitePaper]). Therefore, for operators that are willing to introduce IPv6 the most common approach is to apply the Dual-Stack transition solution, which appears more robust, and easier to troubleshoot and support.

With Dual-Stack, IPv6 can be introduced together with other network upgrades and many parts of network management and IT systems can still work in IPv4. This avoids major upgrade of such systems to support IPv6, which is possibly the most difficult task in the IPv6 transition. In other words, the cost and effort on the network management and IT system upgrade are moderate. The benefits are to start to accommodate future services and save the NAT costs.

The CE has both IPv4 and IPv6 addresses at the WAN side and uses an IPv6 connection to the operator gateway, e.g. Broadband Network Gateway (BNG) or Packet Gateway (PGW) / User Plane Function (UPF). However, the hosts and content servers can still be IPv4 and/or IPv6.

For example, NAT64 can enable IPv6-only hosts to access IPv4 servers. The backbone network underlay can also be IPv4 or IPv6.

Although the Dual-Stack IPv6 transition provides advantages in the IPv6 introductory stage, it does have few disadvantages in the long run, like the duplication of the network resources and states, as well as other limitations for network operation. It also means requiring more IPv4 addresses, so an increase in both Capital Expenses (CAPEX) and Operating Expenses (OPEX). Even if private addresses are being used via Carrier-Grade NAT (CGN), there is extra investment in the CGN devices, logs storage and helpdesk to track CGN-related issues.

For this reason, when IPv4 traffic is vanishingly small or when IPv6 usage increases to more than a given percentage, which highly depends on each network, it could be advantageous to switch to the IPv6-only stage with IPv4aaS. It is difficult to establish the criterion for switching (e.g. to properly identify the upper bound of the IPv4 decrease or the lower bound of the IPv6 increase). In addition to the technical factors, the switch to IPv6-only may also include a loss of customers. Based on operational experience and some measurements of network operators participating in World IPv6 Launch [WIPv6L] where, at June 2021, out of 346 entries 108 exceed 50% of IPv6 traffic volume (31.2%), 72 overcome 60% (20.8%), while 37 go beyond 75% (10.7%), the consensus to move to IPv6-only is when IPv6 traffic volume is between 50% and 60%, which easily happens from since the moment IPv6 is deployed in networks where residential customers traffic is higher than business traffic (because major content providers, as already explained, already use IPv6).

4.2. IPv6-only Service Delivery

The second stage, named IPv6-only and including IPv4 support via IPv4aaS, can be a complex decision that depends on several factors, such as economic aspects, policy and government regulation.

[I-D.ietf-v6ops-transition-comparison] discusses and compares the technical merits of the most common transition solutions for IPv6-only service delivery, 464XLAT [RFC6877], DS-lite [RFC6333], Lightweight 4over6 (lw4o6) [RFC7596], MAP-E [RFC7597], and MAP-T [RFC7599], but without providing an explicit recommendation. As the poll highlights Appendix A, the most widely deployed IPv6 transition solution in the MBB domain is 464XLAT while in the FBB space is DS-Lite.

Both of them are IPv6-only solutions providing IPv4aaS. IPv4aaS offers Dual-Stack service to users and allows an operator to run IPv6-only in the network (typically, the access network). It needs

to be observed that an increasing number of operators, also in the FBB area, tend to prefer 464XLAT over the other transition mechanisms, especially in the case of MBB/FBB convergence.

While it cannot be always the case, IPv6-only transition technologies such as 464XLAT require much less IPv4 public addresses [I-D.ietf-v6ops-transition-comparison], because they make a more efficient usage without restricting the number of ports per subscriber. This contributes to reduce troubleshooting costs and to remove some operational issues related to permanent black-listing of IPv4 address blocks when used via CGN in some services. For example, Sony Play Station Network or OpenDNS imply a higher rotation of IPv4 prefixes in CGN, until they get totally blocked, which means extra CAPEX in new IPv4 transfers.

IPv6-only may be facilitated by the natural upgrade or replacement of CEs because of newer technologies (triple-play, higher bandwidth WAN links, better WiFi technologies, etc.). The CAPEX and OPEX of other parts of the network may be lowered (for example CGN and associated logs) due to the operational simplification of the network.

In terms of addressing needs, for specific applications such as in the case of large mobile operators or large DCs, even the full private address space [RFC1918] is not large enough. Also, Dual-Stack likely leads to duplication of network resources and operations to support both IPv6 and IPv4, which increases the amount of state information in the network. For this reason, in some scenarios (e.g. MBB or DCs) IPv6-only stage could be more efficient from the start since the IPv6 introduction phase.

So, in general, when the Dual-Stack disadvantages outweigh the IPv6-only complexity, it makes sense to apply the transition to IPv6-only. Some network operators already started this process, while others are still waiting.

5. IPv6-only Underlay Network Deployment

IPv6-only alone can be misinterpreted as not supporting IPv4. It can be referred to different portions of the network, to the underlay network, to the overlay network (services), as also mentioned in [I-D.palet-v6ops-ipv6-only].

As opposed to the IPv6-only service delivery (with IPv4aaS) discussed in the previous sections, the IPv6-only network means that the whole network (both operator underlay transport and customer traffic overlay) uses IPv6 as the network protocol for all traffic delivery, but some operators may do IPv6-only at the access network only. This can be accomplished on a case-by-case basis.

As a matter of fact, IPv4 reachability must be provided for a long time to come over IPv6 for IPv6-only endpoints. Most operators are leveraging CGN to extend the life of IPv4 instead of going with IPv4aaS.

When operators (both enterprises and service providers) start to migrate from an IPv4 core, MPLS LDPv4 core, SR-MPLSv4 core to introduce IPv6 in the underlay, they do not necessarily need to dual stack the underlay to maintain both IPv4 and IPv6 address families in the transport layer. Forwarding plane complexity on the Provider (P) core should be kept simple as a single protocol only core. An example could be Software Mesh Framework [RFC5565] which is based on a standard single protocol IPv4-Only or IPv6-Only for core where where IPv4 packets can be tunneled over 4to6 MPLS software over an IPv6-Only core.

Hence, when operators decide to migrate to an IPv6 underlay, the Provider (P) core should be IPv4-only or IPv6-only while Dual-Stack is not the best choice. The underlay could be IPv6-only and allows IPv4 packets to be tunneled using VPN over an IPv6-only core and leveraging Advertising IPv4 Network Layer Routing Information (NLRI) with an IPv6 Next Hop [RFC8950]. Indeed, [RFC8950] specifies the extensions necessary to allow advertising IPv4 NLRI, Virtual Private Network Unicast (VPN-IPv4) NLRI, Multicast Virtual Private Network (MVPN-IPv4) NLRI with a Next Hop address that belongs to the IPv6 protocol. And also, [I-D.ietf-bess-ipv6-only-pe-design] allows dual-stacked functionality without having to dual-stack the interface and without any tunneling mechanisms, resulting in OPEX savings for the elimination of IPv4 addressing and BGP peering. This also enables the quick deployment of IPv6 in a core or Data Center network without provisioning IPv6 links with global unicast address, that can be a long process in very large networks.

Regarding the IPv6 underlay network deployment for Access Network (AN) Metro Edge BNG to NG edge, the current trend is to keep MPLS Data Plane IPv4-only and run IPv4/IPv6 Dual Stack to the Access Network (AN) to Customer RG edge node.

As operators do the transition in the future to IPv6 metro and backbone network, e.g. Segment Routing over IPv6 data plane (SRv6), they are able to start the elimination of IPv4 from the underlay transport network while continuing to provide overlay IPv4 services. Basically, as also showed by the poll among network operators, from a network architecture perspective, it is not recommended to apply Dual-Stack to the transport network per reasons mentioned above about the forwarding plane complexities.

If we consider IPv6-only to mean both operator underlay network and customer VPN traffic, that will take more time. If we look at the long term evolution, IPv6 can bring other advantages like introducing advanced protocols developed only on IPv6.

6. IPv6 Benefits

A comparison of the technical and operational characteristics of IPv4 and IPv6 should quite easily highlight the merits of the latter. On the other hand, the long path IPv6 is still taking to become the dominant network protocol shows that those merits are either not clear enough or not sufficient against other market, managerial or financial reasons. The scope of this section is to list what are generally considered the benefits of IPv6, following discussions at the IETF and other communities. The list below does not mean to be exhaustive; it tries to collect some facts about the use of IPv6.

1. Address space advantage: this is probably the most well-known characteristic of IPv6. The address space is orders of magnitude wider than IPv4's. With the emergence of new digital technologies, such as 5G, IoT and Cloud, new use cases have come into being, posing new requirements to IP networks. Over time, numerous technical and economic stop-gap measures have been developed in an attempt to extend the lifetime of IPv4, but all of these measures add cost and complexity to network infrastructure and raise significant barriers to innovation. It is widely recognized that full transition to IPv6 is the only viable option to ensure future growth and innovation to Internet technology and services. Other sections of this paper have already mentioned the efforts of both network and content providers to evolve their internal infrastructures to be IPv6-only to solve the issue of a constrained address space. From an operational perspective, the wide IPv6 address space avoids the typical renumbering activities that take place either when two networks merge or when an infrastructure needs to span across multiple geographic regions.
2. Government and regulation incentives: the shift to IPv6 is beneficial to create the market conditions to foster innovation and competition. This is, as an example, one of conclusions highlighted in [ARCEP]: "This dearth of IPv4 addresses means that the internet will not stop working but it will stop growing. This shortage is also resulting in a significant increase in the price being charged for these addresses in the secondary market, which creates a real barrier to entry for new entrants to the internet. It has therefore become imperative, for the sake of competition and innovation, that all internet players switch over to IPv6". Governments have a huge responsibility in promoting

IPv6 deployment. Some of them are already adopting policies to encourage IPv6 utilization or enforce increased security on IPv4. So, even without funding the IPv6 transition, governments can recommend to add IPv6 compatibility for every connectivity, service or products bid. This will encourage the network operators and vendors who do not want to miss out on government related bids to evolve their infrastructures to be IPv6 capable. Any public incentives for technical evolution will be bonded to IPv6 capabilities of the technology itself. In this regard, in the United States, the Office of Management and Budget is calling for an implementation plan to have 80% of the IP-enabled resources on Federal networks be IPv6-only by 2025. If resources cannot be converted, then the Federal agency is required to have a plan to retire them. The Call for Comment is at [US-FR] and [US-CIO]. In China, the government launched IPv6 action plan in 2017, which requires that networks, applications and terminal devices will fully support the adoption of IPv6 by the end of 2025 [CN].

3. New functionality and standard: it has been already mentioned the IAB statement [IAB], which asks the IETF, as well as other Standards Developing Organizations (SDOs), to ensure that their standards do not assume IPv4. The IAB expects that the IETF will stop requiring IPv4 compatibility in new or extended protocols. Future IETF protocol work will then optimize for and depend on IPv6. [RFC6540] recommends that all networking standards assume the use of IPv6 and be written so they do not require IPv4. In addition, every RIR worldwide strongly recommends immediate IPv6 adoption. This is an incentive to network equipment vendors to endorse IPv6 and view it as the standards-based solution to the IPv4 address shortage. The introduction of new functionality is one of the technical benefit of IPv6 due to the flexible structure of the packet header. The IETF is currently active in defining operational recommendations for the usage of Extension Headers and Options left open by the base standards.
4. Future Proof: IPv6 was designed from scratch to be future-proof, despite some incompatibilities with IPv4. For example, mechanisms to translate back and forth from IPv4 to IPv6 are already in place in mobile networks worldwide. This allows to introduce newer generations of mobile services, as in the case of 5G applications. While exceptions may exist, IPv6 foresees the absence of NAT along the network path. This allows an operator to remove the technical constraints of handling NAT and the relevant cost. A side effect to that, this approach brings the connectivity model back to the end-to-end client/server paradigm. Worth noting, the vast majority of terminals is IPv6-capable. Content is also greatly available on IPv6. The service

connectivity, from terminal to content, in particular in mobile networks, is reality.

7. Common IPv6 Challenges

There are some areas of improvement, that are often mentioned in the literature and during the discussions on IPv6 deployment. This section highlights these common IPv6 challenges in order to encourage more investigations on these aspects.

7.1. Transition Choices

From an architectural perspective, a service provider or an enterprise may perceive quite a complex task the transition to IPv6, due to the many technical alternatives available and the changes required in management and operations. Moreover, the choice of the method to support the transition may depend on factors specific to the operator's or the enterprise's context, such as the IPv6 network design that fits the service requirements, the deployment strategy, and the service and network operations.

This section briefly highlights the approaches that service providers and enterprises may take and the related challenges.

7.1.1. Service Providers

For fixed operators, the massive software upgrade of CEs to support Dual-Stack already started in most of service provider networks. On average, looking at the global statistics, the IPv6 traffic percentage is currently between 30% and 40% of IPv6. As highlighted earlier, all major content providers have already implemented Dual-Stack access to their services and most of them have implemented IPv6-only in their Data Centers. This aspect could affect the decision on the IPv6 adoption for an operator, but there are also other aspects like the current IPv4 addressing status, CE costs, CGN costs and so on.

Fixed Operators with a Dual-Stack architecture, can start defining and apply a new strategy when reaching the limit in terms of number of IPv4 addresses available. This can be done through CGN or with an IPv6-only approach (IPv4aaS).

On the one hand, most of the fixed operators remain attached to a Dual-Stack architecture and have already employed CGN. In this case it is likely that CGN boosts their ability to supply IPv4 connectivity to CEs for more years to come. On the other hand, only few fixed operators have chosen to move to IPv6-only.

For mobile operators, the situation is quite different since, in some cases, mobile operators are already stretching their IPv4 address space since CGN translation levels have been reached and no more IPv4 public pool addresses are available.

Some mobile operators choose to implement Dual-Stack as first and immediate mitigation solution.

Other mobile operators prefer to move to IPv6-only solution (e.g. 464XLAT) since Dual-Stack only mitigates and does not solve completely the IPv4 address scarcity issue.

For both fixed and mobile operators the approach for the transition is not unique and this brings different challenges in relation to the network architecture and related costs. So each operator needs to do own evaluations for the transition based on the specific situation.

7.1.2. Enterprises and Industrial Internet

At present, the key driver for enterprises relies on upstream service providers. If they run out of IPv4 addresses, it is likely that they start providing native IPv6 and non-native IPv4. So for other networks trying to reach enterprise networks, the IPv6 experience could be better than the transitional IPv4 if the enterprise deploys IPv6 in its public-facing services. IPv6 also shows its advantages in the case of acquisition, indeed when an enterprise merges two networks which use IPv4 private addresses, the address space of the two networks may overlap and this makes the merge difficult. Since several governments are introducing IPv6 policy, all the enterprises providing consulting service to governments are also required to support IPv6 and to show their technical expertise in the IPv6 arena.

Enterprises are shielded from IPv4 address depletion issues due to Enterprises predominantly using Proxy and Non internet routable private [RFC1918], thus do not have the business requirement or technical justification to migrate to IPv6. Enterprises need to find a business case and a strong motivation for IPv6 transition to justify additional CAPEX and OPEX. Also, since ICT is not the core business for most of the enterprises, ICT budget is often constrained and cannot expand considerably. However, there are examples of big enterprises that are considering IPv6 to achieve business targets through a more efficient IPv6 network and to introduce newer services which require future-proof IPv6 network architectures.

Enterprises worldwide, in particular small and medium-sized, are quite late to adopt IPv6, especially on internal networks. In most cases, the enterprise engineers and technicians don't know well how IPv6 works and the problem of application porting to IPv6 looks quite

difficult, even if technically is not a big issue. As highlighted in the relevant poll, the technicians may want to get trained but the management do not see a business need for adoption. This creates an unfortunate cycle where misinformation about the complexity of the IPv6 protocol and unreasonable fears about security and manageability combine with the perceived lack of urgent business needs to prevent adoption of IPv6. In 2019 and 2020, there has been a concerted effort by some grass roots non-profits working with ARIN and APNIC to provide training [ARIN-CG] [ISIF-ASIA-G].

For enterprises, the challenge is that of "First Mover Disadvantage". Compared to network operators that may feel the need of a network evolution towards IPv6, enterprises typically upgrade to new technologies and architectures, such as IPv6, only if it gains them revenue, and this is evident, at least in the short term.

As the most promising protocol for network applications, IPv6 is frequently mentioned in relation to Internet of Things and Industry 4.0. However, its industrial adoption, in particular in smart manufacturing systems, has been much slower than expected. Indeed, as for enterprises, it is important to provide an easy way to familiarize system architects and software developers with the IPv6 protocol.

For Industrial Internet and related IIoT applications, it would be desirable to be able to implement a truly distributed system without dependencies to central components. In this regard the distributed IIoT applications can leverage the configuration-less characteristic of IPv6. In addition, it could be interesting to have the ability to use IP based communication and standard application protocols at every point in the production process and further reduce the use of specialized communication systems.

7.1.3. Cloud and Data Centers

Most CSPs have adopted IPv6 in their internal infrastructure but are also active in gathering IPv4 addresses on the transfer market to serve the current business needs of IPv4 connectivity. As noted in the previous section, most enterprises do not consider the transition to IPv6 as a priority. To this extent, the use of IPv4-based network services by the CSPs will last. Yet, CSPs are struggling to buy IPv4 addresses.

It is interesting to look at how much traffic in a network is going to Caches and Content Delivery Networks (CDNs). The response is expected to be an high percentage, at least higher than 50% in most of the cases. Since all the key Caches and CDNs are IPv6-ready [Cldflr], [Akm], [Ggl], [Ntflx], [Amzn], [Mcrsft], [Vrzn]. So the

percentage of traffic going to the key Caches/CDNs is a good approximation of the potential IPv6 traffic in a network.

The challenge for CSPs is related to the support of non-native IPv4 since most CSPs provide native IPv6. If, in the next years, the scarcity of IPv4 addresses becomes more evident, it is likely that the cost of buying an IPv4 address by a CSP could be charged to their customers.

7.1.4. CEs and user devices

It can be noted that most of the user devices (e.g. smartphones) are already IPv6-enabled since so many years. But there are exceptions, for example, smartTVs and Set-Top Box (STBs) typically had IPv6 support since few years ago, however not all the economies replace them at the same pace.

As already mentioned, ISPs who historically provided public IPv4 addresses to their customers generally still have those IPv4 addresses (unless they chose to transfer them). Some have chosen to put new customers on CGN but without touching existing customers. Because of the extremely small number of customers who notice that IPv4 is done via NAT444, it could be less likely to run out of IPv4 addresses and private IPv4 space. But as IPv4-only devices and traffic reduce, then the need to support private and public IPv4 become less. So the complete support of CEs to IPv6 is also an important challenge and incentive to overcome Dual-Stack towards IPv6-only with IPv4aaS [ANSI].

7.2. Government and Regulators

The global picture shows that the deployment of IPv6 worldwide is not uniform at all [G_stats], [APNIC1]. Countries where either market conditions or local regulators have stimulated the adoption of IPv6 show clear sign of growth.

As an example, zooming into the European Union area, countries such as Belgium, France and Germany are well ahead in terms of IPv6 adoption. The French National Regulator, Arcep, can be considered a good reference of National support to IPv6. [ARCEP] introduced an obligation for the operators awarded with a license to use 5G frequencies (3.4–3.8GHz) in Metropolitan France to be IPv6 compatible. As stated, "the goal is to ensure that services are interoperable and to remove obstacles to using services that are only available in IPv6, as the number of devices in use continues to soar, and because the RIPE NCC has run out of IPv4 addresses". A slow adoption of IPv6 could prevent new Internet services to widespread or create a barrier to entry for newcomers to the market. "IPv6 can help

to increase competition in the telecom industry, and help to industrialize a country for specific vertical sectors".

A renewed industrial policy might be advocated in other countries and regions to stimulate IPv6 adoption. As an example, in the United States, the Office of Management and Budget is also calling for IPv6 adoption [US-FR], [US-CIO]. China is another example of govern supporting a country-wide adoption.

7.3. Network Management and Operations

There are important IPv6 complementary solutions related to Operations, Administration and Maintenance (OAM) that look not so complete compared to IPv4. Network Management System (NMS) has a central role in the modern networks for both network operators and enterprises and its transition is a fundamental challenge. This is because some IPv6 products are not field-proven as for IPv4 even if traditional protocols (e.g. SNMP, RADIUS) already supports IPv6. In addition, incompatible vendor roadmap for the development of new NMS features affects the confidence of network operators or enterprises. For example, YANG is the configuration language for networking but in the real world the data modeling is still vendor dependent.

An important factor is represented by the need for training the network operations workforce. Deploying IPv6 requires it as policies and procedures have to be adjusted in order to successfully plan and complete an IPv6 transition. Staff has to be aware of the best practices for managing IPv4 and IPv6 assets. In addition to network nodes, network management applications and equipment need to be properly configured and in some cases also replaced. This may introduce more complexity and costs for the transition.

7.4. Performance

People tend to compare the performance of IPv6 versus IPv4 to argue or motivate the IPv6 transition. In some cases, IPv6 behaving "worse" than IPv4 may be used as an argument for avoiding the full adoption of IPv6. However, there are some aspects where IPv6 is filling the gap to IPv4. This position is supported when looking at available analytics on two critical parameters: packet loss and latency. These parameters have been constantly monitored over time, but only a few extensive researches and measurement campaigns are currently providing up-to-date information. While performance is undoubtedly an important issue to consider and worth further investigation, reality is that a definitive answer cannot be found on what IP version performs better. Depending on the specific use case and application, IPv6 is better; in others the same applies to IPv4.

7.4.1. IPv6 packet loss and latency

[APNIC5] provides a measurement of both the failure rate and RTT of IPv6 compared against IPv4. Both measures are based on scripts that employs the three-way handshake of TCP. As such, the measurement of the failure rate does not provide a direct measurement of packet loss (that would need an Internet-wide measurement campaign). Said that, despite IPv4 is still performing better, the difference seems to have decreased in recent years. Two reports, namely [RIPE1] and [APRICOT], discussed the associated trend, showing how the average worldwide failure rate of IPv6 is still a bit worse than IPv4. Reasons for this effect may be found in endpoints with an unreachable IPv6 address, routing instability or firewall behavior. Yet, this worsening effect may appear as disturbing for a plain transition to IPv6.

[APNIC5] also compares the latency of both address families. Currently, the worldwide average is still in favor of IPv4. Zooming at the country or even at the operator level, it is possible to get more detailed information and appreciate that cases exist where IPv6 is faster than IPv4. Regions (e.g. Western Europe, Northern America, Southern Asia) and Countries (e.g. US, India, Germany) with an advanced deployment of IPv6 (e.g. >45%) are showing that IPv6 has better performance than IPv4. [APRICOT] highlights how when a difference in performance exists it is often related to asymmetric routing issues. Other possible explanations for a relative latency difference lays on the specificity of the IPv6 header which allows packet fragmentation. In turn, this means that hardware needs to spend cycles to analyze all of the header sections and when it is not capable of handling one of them it drops the packet. Even considering this, a difference in latency stands and sometimes it is perceived as a limiting factor for IPv6. A few measurement campaigns on the behavior of IPv6 in CDNs are also available [MAPRG], [INFOCOM]. The TCP connect time is still higher for IPv6 in both cases, even if the gap has reduced over the analysis time window.

7.4.2. Customer Experience

It is not totally clear if the Customer Experience is in some way perceived as better when IPv6 is used instead of IPv4. In some cases it has been publicly reported by IPv6 content providers, that users have a better experience when using IPv6-only compared to IPv4 [ISOC2]. This could be explained because in the case of an IPv6 user connecting to an application hosted in an IPv6-only Data Centers, the connection is end-to-end, without translations. Instead, when using IPv4 there is a NAT translation either in the CE or in the service provider's network, in addition to IPv4 to IPv6 (and back to IPv4) translation in the IPv6-only content provider Data Center. [ISOC2],

[FB] provide reasons in favor of IPv6. In other cases, the result seems to be still slightly in favor of IPv4 [INFOCOM], [MAPRG], even if the difference between IPv4 and IPv6 tends to vanish over time.

7.5. IPv6 security

Another point that is sometimes considered as a challenge when discussing the transition to IPv6 is related to the Security. [RFC9099] analyzes the operational security issues in several places of a network (enterprises, service providers and residential users). It is also worth considering the additional security issues brought into existence by the applied IPv6 transition technologies used to implement IPv4aaS, e.g. 464XLAT, DS-Lite. Some hints are in the paper [ComputSecur].

The security aspects have to be considered to keep the same level of security as it exists nowadays in an IPv4-only network environment. The autoconfiguration features of IPv6 will require some more attention. Router discovery and address autoconfiguration may produce unexpected results and security holes. The IPsec protocol implementation has initially been set as mandatory in every node of the network, but then relaxed to recommendation due to extremely constrained hardware deployed in some devices e.g., sensors, Internet of Things (IoT).

There are some concerns in terms of the security but, on the other hand, IPv6 offers increased efficiency. There are measurable benefits to IPv6 to notice, like more transparency, improved mobility, and also end to end security (if implemented).

As reported in [ISOC3], comparing IPv6 and IPv4 at the protocol level, one may probably conclude that the increased complexity of IPv6 results in an increased number of attack vectors, that imply more possible ways to perform different types attacks. However, a more interesting and practical question is how IPv6 deployments compare to IPv4 deployments in terms of security. In that sense, there are a number of aspects to consider.

Most security vulnerabilities related to network protocols are based on implementation flaws. Typically, security researchers find vulnerabilities in protocol implementations, which eventually are "patched" to mitigate such vulnerabilities. Over time, this process of finding and patching vulnerabilities results in more robust implementations. For obvious reasons, the IPv4 protocols have benefited from the work of security researchers for much longer, and thus, IPv4 implementations are generally more robust than IPv6. However, this is turning also in the other way around, as with more

IPv6 deployment there may be older IPv4 flaws not discovered or even not resolved anymore by vendors.

Besides the intrinsic properties of the protocols, the security level of the resulting deployments is closely related to the level of expertise of network and security engineers. In that sense, there is obviously much more experience and confidence with deploying and operating IPv4 networks than with deploying and operating IPv6 networks.

Finally, implementation of IPv6 security controls obviously depends on the availability of features in security devices and tools. Whilst there have been improvements in this area, there is a lack of parity in terms of features and/or performance when considering IPv4 and IPv6 support in security devices and tools.

7.5.1. Protocols security issues

It is important to say that IPv6 is not more or less secure than IPv4 and the knowledge of the protocol is the best security measure.

In general there are security concerns related to IPv6 that can be classified as follows:

- o Basic IPv6 protocol (Basic header, Extension Headers, Addressing)
- o IPv6 associated protocols (ICMPv6, NDP, MLD, DNS, DHCPv6)
- o Internet-wide IPv6 security (Filtering, DDoS, Transition Mechanisms)

ICMPv6 is an integral part of IPv6 and performs error reporting and diagnostic functions. Since it is used in many IPv6 related protocols, ICMPv6 packet with multicast address should be filtered carefully to avoid attacks. Neighbor Discovery Protocol (NDP) is a node discovery protocol in IPv6 which replaces and enhances functions of ARP. Multicast Listener Discovery (MLD) is used by IPv6 routers for discovering multicast listeners on a directly attached link, much like Internet Group Management Protocol (IGMP) is used in IPv4.

These IPv6 associated protocols like ICMPv6, NDP and MLD are something new compared to IPv4, so they add new security threats and the related solutions are still under discussion today. NDP has vulnerabilities [RFC3756] [RFC6583]. The specification says to use IPsec but it is impractical and not used, on the other hand, SEND (SEcure Neighbour Discovery) [RFC3971] is not widely available.

[RIPE2] describes the most important threats and solutions regarding IPv6 security.

7.5.2. IPv6 Extension Headers and Fragmentation

IPv6 Extension Headers imply some issues, in particular their flexibility also means an increased complexity, indeed security devices and software must process the full chain of headers while firewalls must be able to filter based on Extension Headers. Additionally, packets with IPv6 Extension Headers may be dropped in the public Internet. Some documents, e.g. [I-D.hinden-6man-hbh-processing], [I-D.bonica-6man-ext-hdr-update], [I-D.peng-v6ops-hbh] analyze and provide guidance regarding the processing procedures of IPv6 Extension Headers.

There are some possible attacks through EHs, for example RH0 can be used for traffic amplification over a remote path and it is deprecated. Other attacks based on Extension Headers are based on IPv6 Header Chains and Fragmentation that could be used to bypass filtering, but to mitigate this effect, Header chain should go only in the first fragment and the use of the IPv6 Fragmentation Header is forbidden in all Neighbor Discovery messages.

Fragment Header is used by IPv6 source node to send a packet bigger than path MTU and the Destination host processes fragment headers. There are several threats related to fragmentation to pay attention to e.g. overlapping fragments (not allowed) resource consumption while waiting for last fragment (to discard), atomic fragments (to be isolated).

A lot of additional functionality has been added to IPv6 primarily by adding Extension Headers and/or using overlay encapsulation. All of these expand the packet size and this could lead to oversized packets that would be dropped on some links. It is important to investigate the potential problems with oversized packets in the first place. Fragmentation must not be done in transit and a better solution needs to be found, e.g. upgrade all links to bigger MTU or follow specific recommendations at the source node. [I-D.vasilenko-v6ops-ipv6-oversized-analysis] analyzes available standards for the resolution of oversized packet drops.

8. Security Considerations

This document has no impact on the security properties of specific IPv6 protocols or transition tools. The security considerations relating to the protocols and transition tools are described in the relevant documents.

9. Contributors

Sebastien Lourdez
Post Luxembourg
Email: sebastien.lourdez@post.lu

10. Acknowledgements

The authors of this document would like to thank Brian Carpenter, Fred Baker, Alexandre Petrescu, Barbara Stark, Haisheng Yu (Johnson), Dhruv Dhody, Gabor Lencse, Shuping Peng, Daniel Voyer, Daniel Bernier, Hariharan Ananthakrishnan, Donavan Fritz, Igor Lubashev, Erik Nygren, Eduard Vasilenko and Xipeng Xiao for their comments and review of this document.

11. IANA Considerations

This document has no actions for IANA.

12. References

12.1. Normative References

- [I-D.ietf-v6ops-transition-comparison]
Lencse, G., Martinez, J. P., Howard, L., Patterson, R., and I. Farrer, "Pros and Cons of IPv6 Transition Technologies for IPv4aaS", draft-ietf-v6ops-transition-comparison-02 (work in progress), March 2022.
- [RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, DOI 10.17487/RFC1918, February 1996, <<https://www.rfc-editor.org/info/rfc1918>>.
- [RFC3756] Nikander, P., Ed., Kempf, J., and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats", RFC 3756, DOI 10.17487/RFC3756, May 2004, <<https://www.rfc-editor.org/info/rfc3756>>.
- [RFC3971] Arkko, J., Ed., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, DOI 10.17487/RFC3971, March 2005, <<https://www.rfc-editor.org/info/rfc3971>>.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, DOI 10.17487/RFC4213, October 2005, <<https://www.rfc-editor.org/info/rfc4213>>.

- [RFC6036] Carpenter, B. and S. Jiang, "Emerging Service Provider Scenarios for IPv6 Deployment", RFC 6036, DOI 10.17487/RFC6036, October 2010, <<https://www.rfc-editor.org/info/rfc6036>>.
- [RFC6180] Arkko, J. and F. Baker, "Guidelines for Using IPv6 Transition Mechanisms during IPv6 Deployment", RFC 6180, DOI 10.17487/RFC6180, May 2011, <<https://www.rfc-editor.org/info/rfc6180>>.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, DOI 10.17487/RFC6333, August 2011, <<https://www.rfc-editor.org/info/rfc6333>>.
- [RFC6540] George, W., Donley, C., Liljenstolpe, C., and L. Howard, "IPv6 Support Required for All IP-Capable Nodes", BCP 177, RFC 6540, DOI 10.17487/RFC6540, April 2012, <<https://www.rfc-editor.org/info/rfc6540>>.
- [RFC6583] Gashinsky, I., Jaeggli, J., and W. Kumari, "Operational Neighbor Discovery Problems", RFC 6583, DOI 10.17487/RFC6583, March 2012, <<https://www.rfc-editor.org/info/rfc6583>>.
- [RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, DOI 10.17487/RFC6877, April 2013, <<https://www.rfc-editor.org/info/rfc6877>>.
- [RFC6883] Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content Providers and Application Service Providers", RFC 6883, DOI 10.17487/RFC6883, March 2013, <<https://www.rfc-editor.org/info/rfc6883>>.
- [RFC7381] Chittimaneni, K., Chown, T., Howard, L., Kuarsingh, V., Pouffary, Y., and E. Vyncke, "Enterprise IPv6 Deployment Guidelines", RFC 7381, DOI 10.17487/RFC7381, October 2014, <<https://www.rfc-editor.org/info/rfc7381>>.
- [RFC7596] Cui, Y., Sun, Q., Boucadair, M., Tsou, T., Lee, Y., and I. Farrer, "Lightweight 4over6: An Extension to the Dual-Stack Lite Architecture", RFC 7596, DOI 10.17487/RFC7596, July 2015, <<https://www.rfc-editor.org/info/rfc7596>>.

- [RFC7597] Troan, O., Ed., Dec, W., Li, X., Bao, C., Matsushima, S., Murakami, T., and T. Taylor, Ed., "Mapping of Address and Port with Encapsulation (MAP-E)", RFC 7597, DOI 10.17487/RFC7597, July 2015, <<https://www.rfc-editor.org/info/rfc7597>>.
- [RFC7599] Li, X., Bao, C., Dec, W., Ed., Troan, O., Matsushima, S., and T. Murakami, "Mapping of Address and Port using Translation (MAP-T)", RFC 7599, DOI 10.17487/RFC7599, July 2015, <<https://www.rfc-editor.org/info/rfc7599>>.
- [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>>.
- [RFC9099] Vyncke, E., Chittimaneni, K., Kaeo, M., and E. Rey, "Operational Security Considerations for IPv6 Networks", RFC 9099, DOI 10.17487/RFC9099, August 2021, <<https://www.rfc-editor.org/info/rfc9099>>.

12.2. Informative References

- [Akm] Akamai, "IPv6 Adaptation", <<https://www.akamai.com/us/en/multimedia/documents/product-brief/ipv6-adaptation-product-brief.pdf>>.
- [Akm-stats] Akamai, "IPv6 Adoption Visualization", 2021, <<https://www.akamai.com/uk/en/resources/our-thinking/state-of-the-internet-report/state-of-the-internet-ipv6-adoption-visualization.jsp>>.
- [Alx] Alexa, "The top 500 sites on the web", 2021, <<https://www.alexa.com/topsites>>.
- [Amzn] Amazon, "Announcing Internet Protocol Version 6 (IPv6) support for Amazon CloudFront, AWS WAF, and Amazon S3 Transfer Acceleration", <<https://aws.amazon.com/es/about-aws/whats-new/2016/10/ipv6-support-for-cloudfront-waf-and-s3-transfer-acceleration/>>.
- [ANSI] ANSI/CTA, "ANSI/CTA Standard Host and Router Profiles for IPv6", 2020, <<https://shop.cta.tech/products/host-and-router-profiles-for-ipv6>>.

- [APNIC1] APNIC, "IPv6 Capable Rate by country (%)", 2022, <<https://stats.labs.apnic.net/ipv6>>.
- [APNIC2] APNIC2, "Addressing 2021", 2022, <<https://blog.apnic.net/2022/01/19/ip-addressing-in-2021/>>.
- [APNIC3] APNIC, "BGP in 2020 - The BGP Table", 2021, <<https://blog.apnic.net/2021/01/05/bgp-in-2020-the-bgp-table/>>.
- [APNIC4] APNIC, "BGP in 2021 - The BGP Table", 2022, <<https://blog.apnic.net/2022/01/06/bgp-in-2021-the-bgp-table/>>.
- [APNIC5] APNIC, "Average RTT Difference (ms) (V6 - V4) for World (XA)", 2022, <<https://stats.labs.apnic.net/v6perf/XA>>.
- [APRICOT] Huston, G., "Average RTT Difference (ms) (V6 - V4) for World (XA)", 2020, <<https://2020.apricot.net/assets/files/APAE432/ipv6-performance-measurement.pdf>>.
- [ARCEP] ARCEP, "Arcep Decision no 2019-1386, Decision on the terms and conditions for awarding licences to use frequencies in the 3.4-3.8GHz band", 2019, <https://www.arcep.fr/uploads/tx_gsavis/19-1386.pdf>.
- [ARIN-CG] ARIN, "Community Grant Program: IPv6 Security, Applications, and Training for Enterprises", 2020, <https://www.arin.net/about/community_grants/recipients/>.
- [ARIN-SW] ARIN, "Preparing Applications for IPv6", <https://www.arin.net/resources/guide/ipv6/preparing_apps_for_v6.pdf>.
- [BGR_1] BIIGROUP, "China Commercial IPv6 and DNSSEC Deployment Monitor", 2022, <<http://218.2.231.237:5001/cgi-bin/generate>>.
- [BGR_2] BIIGROUP, "China Government IPv6 and DNSSEC Deployment Monitor", 2022, <http://218.2.231.237:5001/cgi-bin/generate_gov>.
- [BGR_3] BIIGROUP, "China Education IPv6 and DNSSEC Deployment Monitor", 2022, <http://218.2.231.237:5001/cgi-bin/generate_edu>.

- [CAIDA] APNIC, "Client-Side IPv6 Measurement", 2020, <<https://www.cmand.org/workshops/202006-v6/slides/2020-06-16-client-side.pdf>>.
- [CAIR] Cisco, "Cisco Annual Internet Report (2018-2023) White Paper", 2020, <<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>>.
- [Cldflr] Cloudflare, "Understanding and configuring Cloudflare's IPv6 support", <<https://support.cloudflare.com/hc/en-us/articles/229666767-Understanding-and-configuring-Cloudflare-s-IPv6-support>>.
- [CN] China.org.cn, "China to speed up IPv6-based Internet development", 2017, <http://www.china.org.cn/business/2017-11/27/content_41948814.htm>.
- [CN-IPv6] National IPv6 Deployment and Monitoring Platform, "Active IPv6 Internet users", 2022, <<https://www.china-ipv6.cn/#/activeconnect/simpleInfo>>.
- [CNLABS_1] CNLABS, "Industry IPv6 and DNSSEC Statistics", 2022, <https://cnlabs.in/IPv6_Mon/generate_industry.html>.
- [CNLABS_2] CNLABS, "Industry IPv6 and DNSSEC Statistics", 2022, <https://cnlabs.in/IPv6_Mon/generate_gov.html>.
- [CNLABS_3] CNLABS, "Industry IPv6 and DNSSEC Statistics", 2022, <https://cnlabs.in/IPv6_Mon/generate_industry.html>.
- [ComputSecur] Computers & Security (Elsevier), "Methodology for the identification of potential security issues of different IPv6 transition technologies: Threat analysis of DNS64 and stateful NAT64", DOI 10.1016/j.cose.2018.04.012, 2018.
- [Csc6lab] Cisco, "World - Display Content Data", 2022, <<https://6lab.cisco.com/stats/>>.
- [ETSI-IP6-WhitePaper] ETSI, "ETSI White Paper No. 35: IPv6 Best Practices, Benefits, Transition Challenges and the Way Forward", ISBN 979-10-92620-31-1, 2020.

- [FB] Saab, P., "Facebook IPv6 Experience", 2015,
<<https://youtu.be/An7s25FSK0U>>.
- [G_stats] Google, "Google IPv6 Per-Country IPv6 adoption", 2021,
<<https://www.google.com/intl/en/ipv6/statistics.html#tab=per-country-ipv6-adoption>>.
- [Ggl] Google, "Introduction to GGC",
<<https://support.google.com/interconnect/answer/9058809?hl=en>>.
- [HxBld] HexaBuild, "IPv6 Adoption Report 2020", 2020,
<<https://hexabuild.io/assets/files/HexaBuild-IPv6-Adoption-Report-2020.pdf>>.
- [I-D.bonica-6man-ext-hdr-update]
Bonica, R. and T. Jinmei, "Inserting, Processing And Deleting IPv6 Extension Headers", draft-bonica-6man-ext-hdr-update-07 (work in progress), February 2022.
- [I-D.hinden-6man-hbh-processing]
Hinden, R. M. and G. Fairhurst, "IPv6 Hop-by-Hop Options Processing Procedures", draft-hinden-6man-hbh-processing-01 (work in progress), June 2021.
- [I-D.ietf-bess-ipv6-only-pe-design]
Mishra, G., Mishra, M., Tantsura, J., Madhavi, S., Yang, Q., Simpson, A., and S. Chen, "IPv6-Only PE Design for IPv4-NLRI with IPv6-NH", draft-ietf-bess-ipv6-only-pe-design-00 (work in progress), September 2021.
- [I-D.palet-v6ops-ipv6-only]
Martinez, J. P., "IPv6-only Terminology Definition", draft-palet-v6ops-ipv6-only-05 (work in progress), March 2020.
- [I-D.peng-v6ops-hbh]
Peng, S., Li, Z., Xie, C., Qin, Z., and G. Mishra, "Processing of the Hop-by-Hop Options Header", draft-peng-v6ops-hbh-06 (work in progress), August 2021.
- [I-D.vasilenko-v6ops-ipv6-oversized-analysis]
Vasilenko, E., Xipeng, X., and D. Khaustov, "IPv6 Oversized Packets Analysis", draft-vasilenko-v6ops-ipv6-oversized-analysis-01 (work in progress), September 2021.
- [IAB] IAB, "IAB Statement on IPv6", 2016,
<<https://www.iab.org/2016/11/07/iab-statement-on-ipv6/>>.

- [IGP-GT] Internet Governance Project, Georgia Tech, "The hidden standards war: economic factors affecting IPv6 deployment", 2019, <<https://via.hypothes.is/https://www.internetgovernance.org/wp-content/uploads/IPv6-Migration-Study-final-report.pdf>>.
- [INFOCOM] Doan, T., "A Longitudinal View of Netflix: Content Delivery over IPv6 and Content Cache Deployments", 2020, <<https://dl.acm.org/doi/abs/10.1109/INFOCOM41043.2020.9155367>>.
- [IPv6Forum] IPv6Forum, "Estimating IPv6 & DNSSEC External Service Deployment Status", 2022, <<https://www.ipv6forum.com/IPv6-Monitoring/index.html>>.
- [ISIF-ASIA-G] ISIF Asia, "Internet Operations Research Grant: IPv6 Deployment at Enterprises. IIESoc. India", 2020, <<https://isif.asia/2020-grantees/>>.
- [ISOC1] Internet Society, "State of IPv6 Deployment 2018", 2018, <<https://www.internetsociety.org/resources/2018/state-of-ipv6-deployment-2018/>>.
- [ISOC2] Internet Society, "Facebook News Feeds Load 20-40% Faster Over IPv6", 2015, <<https://www.internetsociety.org/blog/2015/04/facebook-news-feeds-load-20-40-faster-over-ipv6/>>.
- [ISOC3] Internet Society, "IPv6 Security FAQ", 2019, <<https://www.internetsociety.org/wp-content/uploads/2019/02/Deploy360-IPv6-Security-FAQ.pdf>>.
- [MAPRG] Bajpai, V., "Measuring YouTube Content Delivery over IPv6", 2017, <<https://www.ietf.org/proceedings/99/slides/slides-99-maprg-measuring-youtube-content-delivery-over-ipv6-00.pdf>>.
- [Mcrsft] Microsoft, "IPv6 for Azure VMs available in most regions", <<https://azure.microsoft.com/en-us/updates/ipv6-for-azure-vms/>>.
- [NRO] NRO, "Internet Number Resource Status Report", 2021, <<https://www.nro.net/wp-content/uploads/NRO-Statistics-2021-Q3-FINAL.pdf>>.

- [NST_1] NIST, "Estimating Industry IPv6 and DNSSEC External Service Deployment Status", 2022, <<https://fedv6-deployment.antd.nist.gov/cgi-bin/generate-com>>.
- [NST_2] NIST, "Estimating USG IPv6 and DNSSEC External Service Deployment Status", 2022, <<https://fedv6-deployment.antd.nist.gov/cgi-bin/generate-gov>>.
- [NST_3] NIST, "Estimating University IPv6 and DNSSEC External Service Deployment Status", 2022, <<https://fedv6-deployment.antd.nist.gov/cgi-bin/generate-edu>>.
- [Ntflx] Netflix, "Enabling Support for IPv6", <<https://netflixtechblog.com/enabling-support-for-ipv6-48a495d5196f>>.
- [POTAROO1] POTAROO, "IP Addressing through 2021", 2022, <<https://www.potaroo.net/ispcol/2022-01/addr2021.html>>.
- [POTAROO2] POTAROO, "IPv6 Resource Allocations", 2022, <<https://www.potaroo.net/bgp/iso3166/v6cc.html>>.
- [RFC5565] Wu, J., Cui, Y., Metz, C., and E. Rosen, "Software Mesh Framework", RFC 5565, DOI 10.17487/RFC5565, June 2009, <<https://www.rfc-editor.org/info/rfc5565>>.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, DOI 10.17487/RFC6555, April 2012, <<https://www.rfc-editor.org/info/rfc6555>>.
- [RFC8305] Schinazi, D. and T. Pauly, "Happy Eyeballs Version 2: Better Connectivity Using Concurrency", RFC 8305, DOI 10.17487/RFC8305, December 2017, <<https://www.rfc-editor.org/info/rfc8305>>.
- [RIPE1] Huston, G., "Measuring IPv6 Performance", 2016, <<https://ripe73.ripe.net/wp-content/uploads/presentations/35-2016-10-24-v6-performance.pdf>>.
- [RIPE2] RIPE, "IPv6 Security", 2019, <<https://www.ripe.net/support/training/material/ipv6-security/ipv6security-slides.pdf>>.
- [RIPE3] RIPE, "IPv6", 2021, <https://www.ripe.net/participate/meetings/roundtable/26-january-2021/ipv6_roundtable_-jan-2021.pdf>.

- [RlncJ] Reliance Jio, "IPv6-only adoption challenges and standardization requirements", 2020, <<https://datatracker.ietf.org/meeting/109/materials/slides-109-v6ops-ipv6-only-adoption-challenges-and-standardization-requirements-03>>.
- [SNDVN] SANDVINE, "Sandvine releases 2020 Mobile Internet Phenomena Report: YouTube is over 25% of all mobile traffic", 2020, <<https://www.sandvine.com/press-releases/sandvine-releases-2020-mobile-internet-phenomena-report-youtube-is-over-25-of-all-mobile-traffic>>.
- [US-CIO] The CIO Council, "Memorandum for Heads of Executive Departments and Agencies. Completing the Transition to Internet Protocol Version 6 (IPv6)", 2020, <<https://www.cio.gov/assets/resources/internet-protocol-version6-draft.pdf>>.
- [US-FR] Federal Register, "Request for Comments on Updated Guidance for Completing the Transition to the Next Generation Internet Protocol, Internet Protocol Version 6 (IPv6)", 2020, <<https://www.federalregister.gov/documents/2020/03/02/2020-04202/request-for-comments-on-updated-guidance-for-completing-the-transition-to-the-next-generation>>.
- [Vrzn] Verizon, "Verizon Digital Media Services announces IPv6 Compliance", <<https://www.verizondigitalmedia.com/blog/verizon-digital-media-services-announces-ipv6-compliance/>>.
- [W3Tech] W3Tech, "Historical yearly trends in the usage statistics of site elements for websites", 2021, <https://w3techs.com/technologies/history_overview/site_element/all/y>.
- [Wikipedia] Wikipedia, "Comparison of IPv6 support in common applications", <https://en.wikipedia.org/wiki/Comparison_of_IPv6_support_in_common_applications>.
- [WIPv6L] World IPv6 Launch, "World IPv6 Launch - Measurements", 2021, <<https://www.worldipv6launch.org/measurements/>>.

Appendix A. Summary of Questionnaire and Replies for network operators

A survey was proposed to more than 50 service providers in the European region during the third quarter of 2020 to ask for their plans on IPv6 and the status of IPv6 deployment.

40 people, representing 38 organizations, provided a response. This appendix summarizes the results obtained.

Respondents' business

	Convergent	Mobile	Fixed
Type of operators	82%	8%	11%

Question 1. Do you have plan to move more fixed or mobile or enterprise users to IPv6 in the next 2 years?

- a. If so, fixed, or mobile, or enterprise?
- b. What are the reasons to do so?
- c. When to start: already on going, in 12 months, after 12 months?
- d. Which transition solution will you use, Dual-Stack, DS-Lite, 464XLAT, MAP-T/E?

Answer 1.A (38 respondents)

	Yes	No
Plans availability	79%	21%

	Mobile	Fixed	Enterprise	Don't answer
Business segment	63%	63%	50%	3%

Answer 1.B (29 respondents)

Even this was an open question, some common answers can be found.

14 respondents (48%) highlighted issues related to IPv4 depletion. The reason to move to IPv6 is to avoid private and/or overlapping addresses.

For 6 respondents (20%) 5G/IoT is a business incentive to introduce IPv6.

4 respondents (13%) also highlight that there is a National regulation request to enable IPv6 associated with the launch of 5G.

4 respondents (13%) consider IPv6 as a part of their innovation strategy or an enabler for new services.

4 respondents (13%) introduce IPv6 because of Enterprise customers demand.

Answer 1.C (30 respondents)

Timeframe	On-going 60%	In 12 months 33%	After 12 months 0%	Don't answer 7%
-----------	-----------------	---------------------	-----------------------	--------------------

Answer 1.D (28 respondents for cellular, 27 for wireline)

Transition in use Cellular	Dual-Stack 39%	464XLAT 21%	MAP-T 4%	Don't answer 36%
Transition in use Wireline	Dual-Stack 59%	DS-Lite 19%	6RD/6VPE 4%	Don't answer 19%

Question 2. Do you need to change network devices for the above goal?

- a. If yes, what kind of devices: CE, or BNG/mobile core, or NAT?
- b. Will you migrate your metro or backbone or backhaul network to support IPv6?

Answer 2.A (30 respondents)

	Yes	No	Don't answer		
Need of changing	43%	33%	23%		
	CEs	Routers	BNG	CGN	Mobile core
What to change	47%	27%	20%	33%	27%

Answer 2.B (22 respondents)

	Yes	Future	No
Plans for transition	9%	9%	82%

Appendix B. Summary of Questionnaire and Replies for enterprises

The Industry Network Technology Council (INTC) developed the following poll to verify the need or willingness of medium-to-large US-based enterprises for training and consultancy on IPv6 (<https://industry.netcouncil.org/>).

54 organizations provided an answer.

Question 1. How much IPv6 implementation have you done at your organization? (54 respondents)

None	16.67%
Some people have gotten some training	16.67%
Many people have gotten some training	1.85%
Web site is IPv6 enabled	7.41%
Most equipment is dual-stacked	31.48%
Have an IPv6 transition plan for entire network	5.56%
Running native IPv6 in many places	20.37%
Entire network is IPv6-only	0.00%

Question 2. What kind of help or classes would you like to see INTC do? (54 respondents)

Classes/labs on IPv6 security	66.67%
Classes/labs on IPv6 fundamentals	55.56%
Classes/labs on address planning/network conf.	57.41%
Classes/labs on IPv6 troubleshooting	66.67%
Classes/labs on application conversion	35.19%
Other	14.81%

Question 3. As you begin to think about the implementation of IPv6 at your organization, what areas do you feel are of concern? (54 respondents)

Security	31.48%
Application conversion	25.93%
Training	27.78%
All the above	33.33%
Don't know enough to answer	14.81%
Other	9.26%

Authors' Addresses

Giuseppe Fioccola
Huawei Technologies
Riesstrasse, 25
Munich 80992
Germany

Email: giuseppe.fioccola@huawei.com

Paolo Volpato
Huawei Technologies
Via Lorenteggio, 240
Milan 20147
Italy

Email: paolo.volpato@huawei.com

Nalini Elkins
Inside Products
36A Upper Circle
Carmel Valley CA 93924
United States of America

Email: nalini.elkins@insidethestack.com

Jordi Palet Martinez
The IPv6 Company
Molino de la Navata, 75
La Navata - Galapagar, Madrid 28420
Spain

Email: jordi.palet@theipv6company.com

Gyan S. Mishra
Verizon Inc.

Email: gyan.s.mishra@verizon.com

Chongfeng Xie
China Telecom

Email: xiechf@chinatelecom.cn

v6ops
Internet-Draft
Intended status: Informational
Expires: 7 October 2022

G. Lencse
BUTE
J. Palet Martinez
The IPv6 Company
L. Howard
Retevia
R. Patterson
Sky UK
I. Farrer
Deutsche Telekom AG
5 April 2022

Pros and Cons of IPv6 Transition Technologies for IPv4aaS
draft-ietf-v6ops-transition-comparison-03

Abstract

Several IPv6 transition technologies have been developed to provide customers with IPv4-as-a-Service (IPv4aaS) for ISPs with an IPv6-only access and/or core network. All these technologies have their advantages and disadvantages, and depending on existing topology, skills, strategy and other preferences, one of these technologies may be the most appropriate solution for a network operator.

This document examines the five most prominent IPv4aaS technologies considering a number of different aspects to provide network operators with an easy to use reference to assist in selecting the technology that best suits their needs.

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 October 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
2. Overview of the Technologies	4
2.1. 464XLAT	4
2.2. Dual-Stack Lite	5
2.3. Lightweight 4over6	6
2.4. MAP-E	6
2.5. MAP-T	7
3. High-level Architectures and their Consequences	8
3.1. Service Provider Network Traversal	8
3.2. Network Address Translation	9
3.3. IPv4 Address Sharing	10
3.4. IPv4 Pool Size Considerations	11
3.5. CE Provisioning Considerations	13
3.6. Support for Multicast	13
4. Detailed Analysis	14
4.1. Architectural Differences	14
4.1.1. Basic Comparison	14
4.2. Tradeoff between Port Number Efficiency and Stateless Operation	14
4.3. Support for Public Server Operation	17
4.4. Support and Implementations	18
4.4.1. OS Support	18
4.4.2. Support in Cellular and Broadband Networks	18
4.4.3. Implementation Code Sizes	19
4.5. Typical Deployment and Traffic Volume Considerations	19
4.5.1. Deployment Possibilities	19
4.5.2. Cellular Networks with 464XLAT	19
4.5.3. Wireline Networks with 464XLAT	20
4.6. Load Sharing	20
4.7. Logging	21
4.8. Optimization for IPv4-only devices/applications	22
5. Performance Comparison	22

6. Acknowledgements	23
7. IANA Considerations	23
8. Security Considerations	23
9. References	24
9.1. Normative References	24
9.2. Informative References	28
Appendix A. Change Log	30
A.1. 01 - 02	30
A.2. 02 - 03	31
A.3. 03 - 04	31
A.4. 04 - 05	31
A.5. 05 - 06	31
A.6. 06 - 00-WG Item	31
A.7. 00 - 01	31
A.8. 01 - 02	31
A.9. 02 - 03	32
Authors' Addresses	32

1. Introduction

As the deployment of IPv6 becomes more prevalent, it follows that network operators will move to building single-stack IPv6 core and access networks to simplify network planning and operations. However, providing customers with IPv4 services continues to be a requirement for the foreseeable future. To meet this need, the IETF has standardized a number of different IPv4aaS technologies for this [LEN2019] based on differing requirements and deployment scenarios.

The number of technologies that have been developed makes it time consuming for a network operator to identify the most appropriate mechanism for their specific deployment. This document provides a comparative analysis of the most commonly used mechanisms to assist operators with this problem.

Five different IPv4aaS solutions are considered. The following IPv6 transition technologies are covered:

1. 464XLAT [RFC6877]
2. Dual Stack Lite [RFC6333]
3. lw4o6 (Lightweight 4over6) [RFC7596]
4. MAP-E [RFC7597]
5. MAP-T [RFC7599]

We note that [RFC6180] gives guidelines for using IPv6 transition mechanisms during IPv6 deployment addressing a much broader topic, whereas this document focuses on a small part of it.

2. Overview of the Technologies

The following sections introduce the different technologies analyzed in this document, describing some of their most important characteristics.

2.1. 464XLAT

464XLAT may use double translation (stateless NAT46 + stateful NAT64) or single translation (stateful NAT64), depending on different factors, such as the use of DNS by the applications and the availability of a DNS64 function (in the host or in the service provider network).

The customer-side translator (CLAT) is located in the customer's device, and it performs stateless NAT64 translation [RFC7915] (more precisely, stateless NAT46, a stateless IP/ICMP translation from IPv4 to IPv6). IPv4-embedded IPv6 addresses [RFC6052] are used for both source and destination addresses. Commonly, a /96 prefix (either the 64:ff9b::/96 Well-Known Prefix, or a Network-Specific Prefix) is used as the IPv6 destination for the IPv4-embedded client traffic.

In the operator's network, the provider-side translator (PLAT) performs stateful NAT64 [RFC6146] to translate the traffic. The destination IPv4 address is extracted from the IPv4-embedded IPv6 packet destination address and the source address is from a pool of public IPv4 addresses.

Alternatively, when a dedicated /64 is not available for translation, the CLAT device uses a stateful NAT44 translation before the stateless NAT46 translation.

In general, state close to the end-user network (i.e. at the CE - Customer Edge router) is not perceived as problematic as state in the operators network.

In typical deployments, 464XLAT is used together with DNS64 [RFC6147], see Section 3.1.2 of [RFC8683]. When an IPv6-only client or application communicates with an IPv4-only server, the DNS64 server returns the IPv4-embedded IPv6 address of the IPv4-only server. In this case, the IPv6-only client sends out IPv6 packets, and the CLAT functions as an IPv6 router and the PLAT performs a stateful NAT64 for these packets. In this case, there is a single translation.

Alternatively, one can say that DNS64 + stateful NAT64 is used to carry the traffic of the IPv6-only client and the IPv4-only server, and the CLAT is used only for the IPv4 traffic from applications or devices that use literal IPv4 addresses or non-IPv6 compliant APIs.

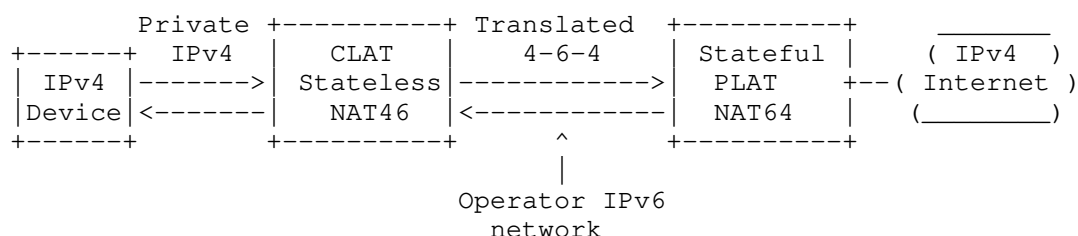


Figure 1: Overview of the 464XLAT architecture

Note: in mobile networks, CLAT is commonly implemented in the user's equipment (UE or smartphone).

2.2. Dual-Stack Lite

Dual-Stack Lite (DS-Lite) [RFC6333] was the first of the considered transition mechanisms to be developed. DS-Lite uses a 'Basic Broadband Bridging' (B4) function in the customer's CE router that encapsulates IPv4 in IPv6 traffic and sends it over the IPv6 native service-provider network to a centralized 'Address Family Transition Router' (AFTR). The AFTR performs encapsulation/decapsulation of the 4in6 [RFC2473] traffic and translates the IPv4 payload to public IPv4 source address using a stateful NAT44 [RFC2663] function.

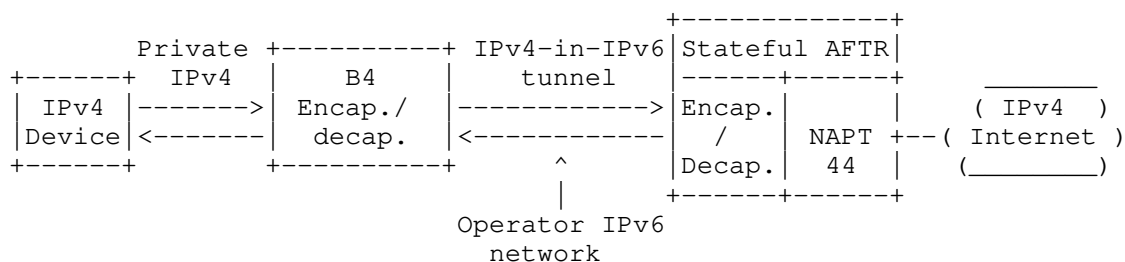


Figure 2: Overview of the DS-Lite architecture

2.3. Lightweight 4over6

Lightweight 4over6 (lw4o6) is a variant of DS-Lite. The main difference is that the stateful NAPT44 function is relocated from the centralized AFTR to the customer's B4 element (called a lwB4). The AFTR (called a lwAFTR) function therefore only performs A+P routing and 4in6 encapsulation/decapsulation.

Routing to the correct client and IPv4 address sharing is achieved using the Address + Port (A+P) model [RFC6346] of provisioning each lwB4 with a unique tuple of IPv4 address and a unique range of layer-4 ports. The client uses these for NAPT44.

The lwAFTR implements a binding table, which has a per-client entry linking the customer's source IPv4 address and allocated range of layer-4 ports to their IPv6 tunnel endpoint address. The binding table allows egress traffic from customers to be validated (to prevent spoofing) and ingress traffic to be correctly encapsulated and forwarded. As there needs to be a per-client entry, an lwAFTR implementation needs to be optimized for performing a per-packet lookup on the binding table.

Direct communication (that is, without translation) between two lwB4s is performed by hair-pinning traffic through the lwAFTR.

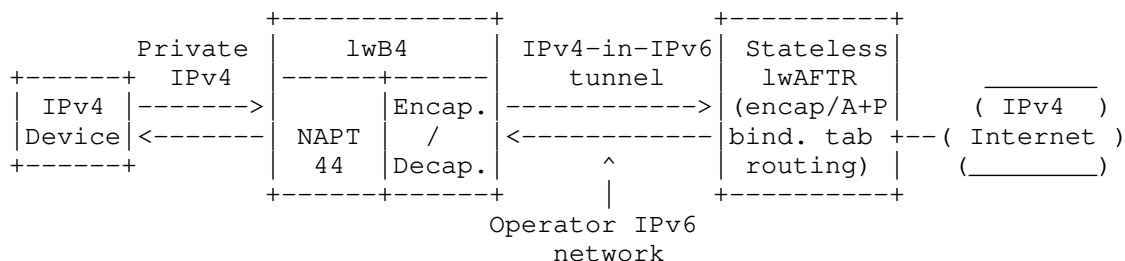


Figure 3: Overview of the lw4o6 architecture

2.4. MAP-E

Like 464XLAT (Section 2.1), MAP-E and MAP-T use [RFC6052] IPv4-embedded IPv6 addresses to represent IPv4 hosts outside the MAP domain.

MAP-E and MAP-T use a stateless algorithm to embed portions of the customer's allocated IPv4 address (or part of an address with A+P routing) into the IPv6 prefix delegated to the client. This allows

for large numbers of clients to be provisioned using a single MAP rule (called a MAP domain). The algorithm also allows for direct IPv4 peer-to-peer communication between hosts provisioned with common MAP rules.

The CE (Customer-Edge) router typically performs stateful NAPT44 [RFC2663] to translate the private IPv4 source addresses and source ports into an address and port range defined by applying the MAP rule to the delegated IPv6 prefix. The client address/port allocation size is a design parameter. The CE router then encapsulates the IPv4 packet in an IPv6 packet [RFC2473] and sends it directly to another host in the MAP domain (for peer-to-peer) or to a Border Router (BR) if the IPv4 destination is not covered in one of the CE's MAP rules.

The MAP BR is provisioned with the set of MAP rules for the MAP domains it serves. These rules determine how the MAP BR is to decapsulate traffic that it receives from client, validating the source IPv4 address and layer 4 ports assigned, as well as how to calculate the destination IPv6 address for ingress IPv4 traffic.

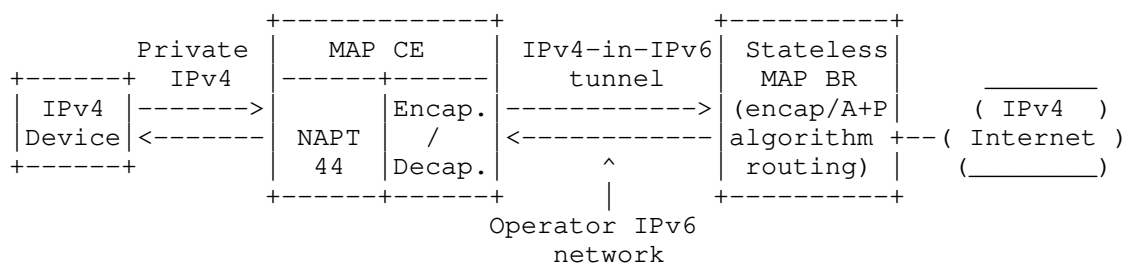


Figure 4: Overview of the MAP-E architecture

2.5. MAP-T

MAP-T uses the same mapping algorithm as MAP-E. The major difference is that double stateless translation (NAT46 in the CE and NAT64 in the BR) is used to traverse the ISP's IPv6 single-stack network. MAP-T can also be compared to 464XLAT when there is a double translation.

A MAP CE router typically performs stateful NAPT44 to translate traffic to a public IPv4 address and port-range calculated by applying the provisioned Basic MAP Rule (BMR - a set of inputs to the algorithm) to the delegated IPv6 prefix. The CE then performs stateless translation from IPv4 to IPv6 [RFC7915]. The MAP BR is provisioned with the same BMR as the client, enabling the received IPv6 traffic to be statelessly NAT64 translated back to the public IPv4 source address used by the client.

Using translation instead of encapsulation also allows IPv4-only nodes to correspond directly with IPv6 nodes in the MAP-T domain that have IPv4-embedded IPv6 addresses.

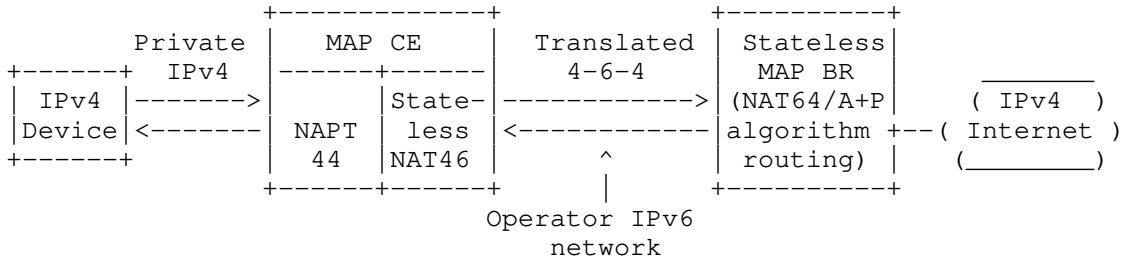


Figure 5: Overview of the MAP-T architecture

3. High-level Architectures and their Consequences

3.1. Service Provider Network Traversal

For the data-plane, there are two approaches for traversing the IPv6 provider network:

- * 4-6-4 translation
- * 4-in-6 encapsulation

	464XLAT	DS-Lite	lw4o6	MAP-E	MAP-T
4-6-4 trans.	X				X
4-6-4 encap.		X	X	X	

Table 1: Available Traversal Mechanisms

In the scope of this document, all of the encapsulation based mechanisms use IP-in-IP tunnelling [RFC2473]. This is a stateless tunneling mechanism which does not require any additional tunnel headers.

It should be noted that both of these approaches result in an increase in the size of the packet that needs to be transported across the operator's network when compared to native IPv4. 4-6-4 translation adds a 20-bytes overhead (the 20-byte IPv4 header is replaced with a 40-byte IPv6 header). Encapsulation has a 40-byte overhead (an IPv6 header is prepended to the IPv4 header).

The increase in packet size can become a significant problem if there is a link with a smaller MTU in the traffic path. This may result in traffic needing to be fragmented at the ingress point to the IPv6 only domain (i.e., the NAT46 or 4in6 encapsulation endpoint). It may also result in the need to implement buffering and fragment re-assembly in the BR node.

The advice given in [RFC7597] Section 8.3.1 is applicable to all of these mechanisms: It is strongly recommended that the MTU in the IPv6-only domain be well managed and that the IPv6 MTU on the CE WAN-side interface be set so that no fragmentation occurs within the boundary of the IPv6-only domain.

3.2. Network Address Translation

For the high-level solution of IPv6 service provider network traversal, MAP-T uses double stateless translation. First at the CE from IPv4 to IPv6 (NAT46), and then from IPv6 to IPv4 (NAT64), at the service provider network.

464XLAT may use double translation (stateless NAT46 + stateful NAT64) or single translation (stateful NAT64), depending on different factors, such as the use of DNS by the applications and the availability of a DNS64 function (in the host or in the service provider network). For deployment guidelines, please refer to [RFC8683].

The first step for the double translation mechanisms is a stateless NAT from IPv4 to IPv6 implemented as SIIT (Stateless IP/ICMP Translation Algorithm) [RFC7915], which does not translate IPv4 header options and/or multicast IP/ICMP packets. With encapsulation-based technologies the header is transported intact and multicast can also be carried.

Single and double translation results in native IPv6 traffic with a layer-4 next-header. The fields in these headers can be used for functions such as hashing across equal-cost multipaths or ACLs. For encapsulation, there is an IPv6 header followed by an IPv4 header. This results in less entropy for hashing algorithms, and may mean that devices in the traffic path that perform header inspection (e.g. router ACLs or firewalls) require the functionality to look into the payload header.

Solutions using double translation can only carry port-aware IP protocols (e.g. TCP, UDP) and ICMP when they are used with IPv4 address sharing (please refer to Section 4.3 for more details). Encapsulation based solutions can carry any other protocols over IP, too.

An in-depth analysis of stateful NAT64 can be found in [RFC6889].

3.3. IPv4 Address Sharing

As public IPv4 address exhaustion is a common motivation for deploying IPv6, transition technologies need to provide a solution for allowing public IPv4 address sharing.

In order to fulfill this requirement, a stateful NAPT function is a necessary function in all of the mechanisms. The major differentiator is where in the architecture this function is located.

The solutions compared by this document fall into two categories:

- * CGN-based approaches (DS-Lite, 464XLAT)
- * A+P-based approaches (lw4o6, MAP-E, MAP-T)

In the CGN-based model, a device such as a CGN/AFTR or NAT64 performs the NAPT44 function and maintains per-session state for all of the active client's traffic. The customer's device does not require per-session state for NAPT.

In the A+P-based model, a device (usually a CE) performs stateful NAPT44 and maintains per-session state only co-located devices, e.g. in the customer's home network. Here, the centralized network function (lwAFTR or BR) only needs to perform stateless encapsulation/decapsulation or NAT64.

Issues related to IPv4 address sharing mechanisms are described in [RFC6269] and should also be considered.

The address sharing efficiency of the five technologies is significantly different, it is discussed in Section 4.2.

lw4o6, MAP-E and MAP-T can also be configured without IPv4 address sharing, see the details in Section 4.3. However, in that case, there is no advantage in terms of public IPv4 address saving. In the case of 464XLAT, this can be achieved as well through EAMT [RFC7757].

Conversely, both MAP-E and MAP-T may be configured to provide more than one public IPv4 address (i.e., an IPv4 prefix shorter than a /32) to customers.

Dynamic DNS issues in address-sharing contexts and their possible solutions using PCP (Port Control Protocol) are discussed in detail in [RFC7393].

3.4. IPv4 Pool Size Considerations

In most networks, it is possible to, using existing data about flows to CDNs/caches or other well-known IPv6-enabled destinations, calculate the percentage of traffic that would turn into IPv6 if it is enabled on that network or part of it.

Knowing that, it is possible to calculate the IPv4 pool size required for a given number of subscribers, depending on the IPv4aaS technology being used.

Often it is assumed that each user-device (computer, tablet, smartphone) behind a NAT, could simultaneously use about 300 ports. Typically, in the case of a residential subscriber, there will be a maximum of 4 of those devices in use simultaneously, which means a total of 1,200 ports.

If for example, 80% of the traffic is expected towards IPv6 destinations, only 20% will actually be using IPv4 ports, so in our example, that will mean 240 ports required per subscriber.

From the 65,535 ports available per IPv4 address, we could even consider reserving 1,024 ports, in order to allow customers that need EAMT entries for incoming connections to System Ports (0-1023, also called well-known ports) [RFC7605], which means 64,511 ports actually available per each IPv4 address.

According to this, a /22 (1.024 public IPv4 addresses) will be sufficient for over 275,000 subscribers
($1,024 \times 64,511 / 240 = 275,246.93$).

Similarly, a /18 (16,384 public IPv4 addresses) will be sufficient for over 4,403,940 subscribers, and so on.

This is a conservative approach, which is valid in the case of 464XLAT, because ports are assigned dynamically by the NAT64, so it is not necessary to consider if one user is actually using more or less ports: Average values work well.

As the deployment of IPv6 progresses, the use of NAT64, and therefore of public IPv4 addresses, decreases (more IPv6/ports, less IPv4/ports), so either more subscribers can be accommodated with the same number of IPv4 addresses, or some of those addressed can be retired from the NAT64.

For comparison, if dual-stack is being used, any given number of users will require the same number of public IPv4 addresses. For instance, a /14 will provide 262,144 IPv4 public addresses for 262,144 subscribers, versus 275,000 subscribers being served with a only a /22.

In the other IPv4aaS technologies, this calculation will only match if the assignment of ports per subscriber can be done dynamically, which is not always the case (depending on the vendor implementation).

An alternative approximation for the other IPv4aaS technologies, when dynamically assignment of addresses is not possible, must ensure sufficient number of ports per subscriber. That means 1,200 ports, and typically, it comes to 2,000 ports in many deployments. In that case, assuming 80% of IPv6 traffic, as above, which will allow only 30 subscribers per each IPv4 address, so the closer approximation to 275,000 subscribers per our example with 464XLAT (with a /22), will be using a /19, which serves 245,760 subscribers (a /19 has 8,192 addresses, 30 subscribers with 2,000 ports each, per address).

If the CGN (in case of DS-Lite) or the CE (in case of lw4o6, MAP-E and MAP-T) make use of a 5-tuple for tracking the NAT connections, the number of ports required per subscriber can be limited as low as 4 ports per subscriber. However, the practical limit depends on the desired limit for parallel connections that any single host behind the NAT can have to the same address and port in Internet. Note that it is becoming more common that applications use AJAX and similar mechanisms, so taking that extreme limit is probably not a very a safe choice.

This extremely reduced number of ports "feature" could also be used in case the CLAT-enabled CE with 464XLAT makes use of the 5-tuple NAT connections tracking, and could also be further extended if the NAT64 also use the 5-tuple.

3.5. CE Provisioning Considerations

All of the technologies require some provisioning of customer devices. The table below shows which methods currently have extensions for provisioning the different mechanisms.

Provisioning Method	464XLAT	DS-Lite	lw4o6	MAP-E	MAP-T
DHCPv6 [RFC8415]		X	X	X	X
RADIUS [RFC8658]		[RFC6519]	X	X	X
TR-069	*	X	*	X	X
DNS64 [RFC7050]	X				
YANG [RFC7950]	[RFC8512]	X	X	X	X
DHCP4o6 [RFC7341]			X	X	

Table 2: Available Provisioning Mechanisms

*: Work started at BroadBand Forum (2021).

X: Supported by the provisioning method.

3.6. Support for Multicast

The solutions covered in this document are all intended for unicast traffic. [RFC8114] describes a method for carrying encapsulated IPv4 multicast traffic over an IPv6 multicast network. This could be deployed in parallel to any of the operator's chosen IPv4aaS mechanism.

4. Detailed Analysis

4.1. Architectural Differences

4.1.1. Basic Comparison

The five IPv4aaS technologies can be classified into $2 \times 2 = 4$ categories on the basis of two aspects:

- * Technology used for service provider network traversal. It can be single/double translation or encapsulation.
- * Presence or absence of NAT44 per-flow state in the operator network.

	464XLAT	DS-Lite	lw4o6	MAP-E	MAP-T
4-6-4 trans.	X				X
4-in-4 encap.		X	X	X	
Per-flow state in op. network	X	X			

Table 3: Available Provisioning Mechanisms

4.2. Tradeoff between Port Number Efficiency and Stateless Operation

464XLAT and DS-Lite use stateful NAT at the PLAT/AFTR devices, respectively. This may cause scalability issues for the number of clients or volume of traffic, but does not impose a limitation on the number of ports per user, as they can be allocated dynamically on-demand and the allocation policy can be centrally managed/adjusted.

A+P based mechanisms (lw4o6, MAP-E, and MAP-T) avoid using NAT in the service provider network. However, this means that the number of ports provided to each user (and hence the effective IPv4 address sharing ratio) must be pre-provisioned to the client.

Changing the allocated port ranges with A+P based technologies, requires more planning and is likely to involve re-provisioning both hosts and operator side equipment. It should be noted that due to the per-customer binding table entry used by lw4o6, a single customer can be re-provisioned (e.g., if they request a full IPv4 address) without needing to change parameters for a number of customers as in a MAP domain.

It is also worth noting that there is a direct relationship between the efficiency of customer public port-allocations and the corresponding logging overhead that may be necessary to meet data-retention requirements. This is considered in Section 4.7 below.

Determining the optimal number of ports for a fixed port set is not an easy task, and may also be impacted by local regulatory law, which may define a maximum number of users per IP address, and consequently a minimum number of ports per user.

On the one hand, the "lack of ports" situation may cause serious problems in the operation of certain applications. For example, Miyakawa has demonstrated the consequences of the session number limitation due to port number shortage on the example of Google Maps [MIY2010]. When the limit was 15, several blocks of the map were missing, and the map was unusable. This study also provided several examples for the session numbers of different applications (the highest one was Apple's iTunes: 230-270 ports).

The port number consumption of different applications is highly varying and e.g. in the case of web browsing it depends on several factors, including the choice of the web page, the web browser, and sometimes even the operating system [REP2014]. For example, under certain conditions, 120-160 ports were used (URL: sohu.com, browser: Firefox under Ubuntu Linux), and in some other cases it was only 3-12 ports (URL: twitter.com, browser: Iceweasel under Debian Linux).

There may be several users behind a CE router, especially in the broadband case (e.g. Internet is used by different members of a family simultaneously), so sufficient ports must be allocated to avoid impacting user experience.

Furthermore, assigning too many ports per CE router will result in waste of public IPv4 addresses, which is a scarce and expensive resource. Clearly this is a big advantage in the case of 464XLAT where they are dynamically managed, so that the number of IPv4 addresses for the sharing-pool is smaller while the availability of ports per user don't need to be pre-defined and is not a limitation for them.

There is a direct tradeoff between the optimization of client port allocations and the associated logging overhead. Section 4.7 discusses this in more depth.

We note that common CE router NAT44 implementations utilizing Netfilter, multiplexes active sessions using a 3-tuple (source address, destination address, and destination port). This means that external source ports can be reused for unique internal source and destination address and port sessions. It is also noted, that Netfilter cannot currently make use of multiple source port ranges (i.e. several blocks of ports distributed across the total port space as is common in MAP deployments), this may influence the design when using stateless technologies.

Stateful technologies, 464XLAT and DS-Lite (and also NAT444) can therefore be much more efficient in terms of port allocation and thus public IP address saving. The price is the stateful operation in the service provider network, which allegedly does not scale up well. It should be noticed that in many cases, all those factors may depend on how it is actually implemented.

Measurements have been started to examine the scalability of a few stateful solutions in two areas:

- * How their performance scales up with the number of CPU cores?
- * To what extent their performance degrades with the number of concurrent connections?

The details of the measurements and their results are available from [I-D.lencse-v6ops-transition-scalability].

We note that some CGN-type solutions can allocate ports dynamically "on the fly". Depending on configuration, this can result in the same customer being allocated ports from different source addresses. This can cause operational issues for protocols and applications that expect multiple flows to be sourced from the same address. E.g., ECMP hashing, STUN, gaming, content delivery networks. However, it should be noticed that this is the same problem when a network has a NAT44 with multiple public IPv4 addresses, or even when applications in a dual-stack case, behave wrongly if happy eyeballs is flapping the flow address between IPv4 and IPv6.

The consequences of IPv4 address sharing [RFC6269] may impact all five technologies. However, when ports are allocated statically, more customers may get ports from the same public IPv4 address, which may result in negative consequences with higher probability, e.g. many applications and service providers (Sony PlayStation Network, OpenDNS, etc.) permanently blocking IPv4 ranges if they detect that they are used for address sharing.

Both cases are, again, implementation dependent.

We note that although it is not of typical use, one can do deterministic, stateful NAT and reserve a fixed set of ports for each customer, as well.

4.3. Support for Public Server Operation

Mechanisms that rely on operator side per-flow state do not, by themselves, offer a way for customers to present services on publicly accessible layer-4 ports.

Port Control Protocol (PCP) [RFC6887] provides a mechanism for a client to request an external public port from a CGN device. For server operation, it is required with NAT64/464XLAT, and it is supported in some DS-Lite AFTR implementations.

A+P based mechanisms distribute a public IPv4 address and restricted range of layer-4 ports to the client. In this case, it is possible for the user to configure their device to offer a publicly accessible server on one of their allocated ports. It should be noted that commonly operators do not assign the Well-Known-Ports to users (unless they are allocating a full IPv4 address), so the user will need to run the service on an allocated port, or configure port translation.

Lw4o6, MAP-E and MAP-T may be configured to allocated clients with a full IPv4 address, allowing exclusive use of all ports, and non-port-based layer 4 protocols. Thus, they may also be used to support server/services operation on their default ports. However, when public IPv4 addresses are assigned to the CE router without address sharing, obviously there is no advantage in terms of IPv4 public addresses saving.

It is also possible to configure specific ports mapping in 464XLAT/NAT64 using EAMT [RFC7757], which means that only those ports are "lost" from the pool of addresses, so there is a higher maximization of the total usage of IPv4/port resources.

4.4. Support and Implementations

4.4.1. OS Support

A 464XLAT client (CLAT) is implemented in Windows 10, Linux (including Android), Windows Mobile, Chrome OS and iOS, but at the time of writing is not available in MacOS.

The remaining four solutions are commonly deployed as functions in the CE device only, however in general, except DS-Lite, the vendors support is poor.

The OpenWRT Linux based open-source OS designed for CE devices offers a number of different 'opkg' packages as part of the distribution:

- * '464xlat' enables support for 464XLAT CLAT functionality
- * 'ds-lite' enables support for DSLite B4 functionality
- * 'map' enables support for MAP-E and lw4o6 CE functionality
- * 'map-t' enables support for MAP-T CE functionality

At the time of publication some free open-source implementations exist for the operator side functionality:

- * Jool [jool] (CLAT, NAT64, EAMT, MAP-T CE, MAP-T BR).
- * VPP/fd.io [vpp] (MAP-BR, lwAFTR, CGN, CLAT, NAT64).
- * Snabb [snabb] (lwAFTR).
- * AFTR [aftr] (DSLite AFTR).

4.4.2. Support in Cellular and Broadband Networks

Several cellular networks use 464XLAT, whereas there are no deployments of the four other technologies in cellular networks, as they are neither standardised nor implemented in UE devices.

In broadband networks, there are some deployments of 464XLAT, MAP-E and MAP-T. Lw4o6 and DS-Lite have more deployments, with DS-Lite being the most common, but lw4o6 taking over in the last years.

Please refer to Table 2 and Table 3 of [LEN2019] for a limited set of deployment information.

4.4.3. Implementation Code Sizes

As hint to the relative complexity of the mechanisms, the following code sizes are reported from the OpenWRT implementations of each technology are 17kB, 35kB, 15kB, 35kB, and 48kB for 464XLAT, lw4o6, DS-Lite, MAP-E, MAP-T, and lw4o6, respectively (<https://openwrt.org/packages/start>).

We note that the support for all five technologies requires much less code size than the total sum of the above quantities, because they contain a lot of common functions (data plane is shared among several of them).

4.5. Typical Deployment and Traffic Volume Considerations

4.5.1. Deployment Possibilities

Theoretically, all five IPv4aaS technologies could be used together with DNS64 + stateful NAT64, as it is done in 464XLAT. In this case the CE router would treat the traffic between an IPv6-only client and IPv4-only server as normal IPv6 traffic, and the stateful NAT64 gateway would do a single translation, thus offloading this kind of traffic from the IPv4aaS technology. The cost of this solution would be the need for deploying also DNS64 + stateful NAT64.

However, this has not been implemented in clients or actual deployments, so only 464XLAT always uses this optimization and the other four solutions do not use it at all.

4.5.2. Cellular Networks with 464XLAT

Figures from existing deployments (end of 2018), show that the typical traffic volumes in an IPv6-only cellular network, when 464XLAT technology is used together with DNS64, are:

- * 75% of traffic is IPv6 end-to-end (no translation)
- * 24% of traffic uses DNS64 + NAT64 (1 translation)
- * Less than 1% of traffic uses the CLAT in addition to NAT64 (2 translations), due to an IPv4 socket and/or IPv4 literal.

Without using DNS64, 25% of the traffic would undergo double translation.

4.5.3. Wireline Networks with 464XLAT

Figures from several existing deployments (end of 2020), mainly with residential customers, show that the typical traffic volumes in an IPv6-only network, when 464XLAT is used with DNS64, are in the following ranges:

- * 65%-85% of traffic is IPv6 end-to-end (no translation)
- * 14%-34% of traffic uses DNS64 + NAT64 (1 translation)
- * Less than 1-2% of traffic uses the CLAT in addition to NAT64 (2 translations), due to an IPv4 socket and/or IPv4 literal.

Without using DNS64, 16%-35% of the traffic would undergo double translation.

4.6. Load Sharing

If multiple network-side devices are needed as PLAT/AFTR/BR for capacity, then there is a need for a load sharing mechanism. ECMP (Equal-Cost Multi-Path) load sharing can be used for all technologies, however stateful technologies will be impacted by changes in network topology or device failure.

Technologies utilizing DNS64 can also distribute load across PLAT/AFTR devices, evenly or unevenly, by using different prefixes. Different network specific prefixes can be distributed for subscribers in appropriately sized segments (like split-horizon DNS, also called DNS views).

Stateless technologies, due to the lack of per-flow state, can make use of anycast routing for load sharing and resiliency across network-devices, both ingress and egress; flows can take asymmetric paths through the network, i.e., in through one lwAFTR/BR and out via another.

Mechanisms with centralized NAPT44 state have a number of challenges specifically related to scaling and resilience. As the total amount of client traffic exceeds the capacity of a single CGN instance, additional nodes are required to handle the load. As each CGN maintains a stateful table of active client sessions, this table may need to be synchronized between CGN instances. This is necessary for two reasons:

- * To prevent all active customer sessions being dropped in event of a CGN node failure.

- * To ensure a matching state table entry for an active session in the event of asymmetric routing through different egress and ingress CGN nodes.

4.7. Logging

In the case of 464XLAT and DS-Lite, the user of any given public IPv4 address and port combination will vary over time, therefore, logging is necessary to meet data retention laws. Each entry in the PLAT/AFTR's generates a logging entry. As discussed in Section 4.2, a client may open hundreds of sessions during common tasks such as web-browsing, each of which needs to be logged so the overall logging burden on the network operator is significant. In some countries, this level of logging is required to comply with data retention legislation.

One common optimization available to reduce the logging overhead is the allocation of a block of ports to a client for the duration of their session. This means that logging entry only needs to be made when the client's port block is released, which dramatically reducing the logging overhead. This comes at the cost of less efficient public address sharing as clients need to be allocated a port block of a fixed size regardless of the actual number of ports that they are using.

Stateless technologies that pre-allocate the IPv4 addresses and ports only require that copies of the active MAP rules (for MAP-E and MAP-T), or binding-table (for lw4o6) are retained along with timestamp information of when they have been active. Support tools (e.g., those used to serve data retention requests) may need to be updated to be aware of the mechanism in use (e.g., implementing the MAP algorithm so that IPv4 information can be linked to the IPv6 prefix delegated to a client). As stateless technologies do not have a centralized stateful element which customer traffic needs to pass through, so if data retention laws mandate per-session logging, there is no simple way of meeting this requirement with a stateless technology alone. Thus a centralized NAT44 model may be the only way to meet this requirement.

Deterministic CGN [RFC7422] was proposed as a solution to reduce the resource consumption of logging.

4.8. Optimization for IPv4-only devices/applications

When IPv4-only devices or applications are behind a CE connected with IPv6-only and IPv4aaS, the IPv4-only traffic flows will necessarily, be encapsulated/decapsulated (in the case of DS-Lite, lw4o6 and MAP-E) and will reach the IPv4 address of the destination, even if that service supports dual-stack. This means that the traffic flow will cross thru the AFTR, lwAFTR or BR, depending on the specific transition mechanism being used.

Even if those services are directly connected to the operator network (for example, CDNs, caches), or located internally (such as VoIP, etc.), it is not possible to avoid that overhead.

However, in the case of those mechanism that use a NAT46 function, in the CE (464XLAT and MAP-T), it is possible to take advantage of optimization functionalities, such as the ones described in [I-D.ietf-v6ops-464xlat-optimization].

Using those optimizations, because the NAT46 has already translated the IPv4-only flow to IPv6, and the services are dual-stack, they can be reached without the need to translate them back to IPv4.

5. Performance Comparison

We plan to compare the performances of the most prominent free software implementations of the five IPv6 transition technologies using the methodology described in "Benchmarking Methodology for IPv6 Transition Technologies" [RFC8219].

The Dual DUT Setup of [RFC8219] makes it possible to use the existing "Benchmarking Methodology for Network Interconnect Devices" [RFC2544] compliant measurement devices, however, this solution has two kinds of limitations:

- * Dual DUT setup has the drawback that the performances of the CE and of the ISP side device (e.g. the CLAT and the PLAT of 464XLAT) are measured together. In order to measure the performance of only one of them, we need to ensure that the desired one is the bottleneck.
- * Measurements procedures for PDV and IPDV measurements are missing from the legacy devices, and the old measurement procedure for Latency has been redefined in [RFC8219].

The Single DUT Setup of [RFC8219] makes it possible to benchmark the selected device separately, but it either requires a special Tester or some trick is need, if we want to use legacy Testers. An example

for the latter is our stateless NAT64 measurements testing Throughput and Frame Loss Rate using a legacy [RFC5180] compliant commercial tester [LEN2020a]

Siitperf, an [RFC8219] compliant DPDK-based software Tester for benchmarking stateless NAT64 gateways has been developed recently and it is available from GitHub [SIITperf] as free software and documented in [LEN2021]. Originally, it literally followed the test frame format of [RFC2544] including "hard wired" source and destination port numbers, and then it has been complemented with the random port feature required by [RFC4814]. The new version is documented in [LEN2020b]

Further DPDK-based, [RFC8219] compliant software testers are being developed at the Budapest University of Technology and Economics as student projects. They are planned to be released as free software, too.

Information about the benchmarking tools, measurements and results will be made available in [I-D.lencse-v6ops-transition-benchmarking].

6. Acknowledgements

The authors would like to thank Ole Troan and Warren Kumari for their thorough review of this draft and acknowledge the inputs of Mark Andrews, Edwin Cordeiro, Fred Baker, Alexandre Petrescu, Cameron Byrne, Tore Anderson, Mikael Abrahamsson, Gert Doering, Satoru Matsushima, Mohamed Boucadair, Tom Petch, Yannis Nikolopoulos, and Havard Eidnes.

7. IANA Considerations

This document does not make any request to IANA.

8. Security Considerations

According to the simplest model, the number of bugs is proportional to the number of code lines. Please refer to Section 4.4.3 for code sizes of CE implementations.

For all five technologies, the CE device typically contains a DNS proxy. However, the user may change DNS settings. If it happens and lw4o6, MAP-E and MAP-T are used with significantly restricted port set, which is required for an efficient public IPv4 address sharing, the entropy of the source ports is significantly lowered (e.g. from 16 bits to 10 bits, when 1024 port numbers are assigned to each subscriber) and thus these technologies are theoretically less resilient against cache poisoning, see [RFC5452]. However, an

efficient cache poisoning attack requires that the subscriber operates an own caching DNS server and the attack is performed in the service provider network. Thus, we consider the chance of the successful exploitation of this vulnerability as low.

An in-depth security analysis of all five IPv6 transition technologies and their most prominent free software implementations according to the methodology defined in [LEN2018] is planned.

As the first step, an initial security analysis of 464XLAT was done in [Azz2021].

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>>.
- [RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, DOI 10.17487/RFC2473, December 1998, <<https://www.rfc-editor.org/info/rfc2473>>.
- [RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, DOI 10.17487/RFC2544, March 1999, <<https://www.rfc-editor.org/info/rfc2544>>.
- [RFC2663] Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", RFC 2663, DOI 10.17487/RFC2663, August 1999, <<https://www.rfc-editor.org/info/rfc2663>>.
- [RFC4814] Newman, D. and T. Player, "Hash and Stuffing: Overlooked Factors in Network Device Benchmarking", RFC 4814, DOI 10.17487/RFC4814, March 2007, <<https://www.rfc-editor.org/info/rfc4814>>.
- [RFC5180] Popoviciu, C., Hamza, A., Van de Velde, G., and D. Dugatkin, "IPv6 Benchmarking Methodology for Network Interconnect Devices", RFC 5180, DOI 10.17487/RFC5180, May 2008, <<https://www.rfc-editor.org/info/rfc5180>>.

- [RFC5452] Hubert, A. and R. van Mook, "Measures for Making DNS More Resilient against Forged Answers", RFC 5452, DOI 10.17487/RFC5452, January 2009, <<https://www.rfc-editor.org/info/rfc5452>>.
- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, DOI 10.17487/RFC6052, October 2010, <<https://www.rfc-editor.org/info/rfc6052>>.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, DOI 10.17487/RFC6146, April 2011, <<https://www.rfc-editor.org/info/rfc6146>>.
- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, DOI 10.17487/RFC6147, April 2011, <<https://www.rfc-editor.org/info/rfc6147>>.
- [RFC6180] Arkko, J. and F. Baker, "Guidelines for Using IPv6 Transition Mechanisms during IPv6 Deployment", RFC 6180, DOI 10.17487/RFC6180, May 2011, <<https://www.rfc-editor.org/info/rfc6180>>.
- [RFC6269] Ford, M., Ed., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269, DOI 10.17487/RFC6269, June 2011, <<https://www.rfc-editor.org/info/rfc6269>>.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, DOI 10.17487/RFC6333, August 2011, <<https://www.rfc-editor.org/info/rfc6333>>.
- [RFC6346] Bush, R., Ed., "The Address plus Port (A+P) Approach to the IPv4 Address Shortage", RFC 6346, DOI 10.17487/RFC6346, August 2011, <<https://www.rfc-editor.org/info/rfc6346>>.
- [RFC6519] Maglione, R. and A. Durand, "RADIUS Extensions for Dual-Stack Lite", RFC 6519, DOI 10.17487/RFC6519, February 2012, <<https://www.rfc-editor.org/info/rfc6519>>.

- [RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, DOI 10.17487/RFC6877, April 2013, <<https://www.rfc-editor.org/info/rfc6877>>.
- [RFC6887] Wing, D., Ed., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, DOI 10.17487/RFC6887, April 2013, <<https://www.rfc-editor.org/info/rfc6887>>.
- [RFC6889] Penno, R., Saxena, T., Boucadair, M., and S. Sivakumar, "Analysis of Stateful 64 Translation", RFC 6889, DOI 10.17487/RFC6889, April 2013, <<https://www.rfc-editor.org/info/rfc6889>>.
- [RFC7050] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", RFC 7050, DOI 10.17487/RFC7050, November 2013, <<https://www.rfc-editor.org/info/rfc7050>>.
- [RFC7341] Sun, Q., Cui, Y., Siodelski, M., Krishnan, S., and I. Farrer, "DHCPv4-over-DHCPv6 (DHCP 4o6) Transport", RFC 7341, DOI 10.17487/RFC7341, August 2014, <<https://www.rfc-editor.org/info/rfc7341>>.
- [RFC7393] Deng, X., Boucadair, M., Zhao, Q., Huang, J., and C. Zhou, "Using the Port Control Protocol (PCP) to Update Dynamic DNS", RFC 7393, DOI 10.17487/RFC7393, November 2014, <<https://www.rfc-editor.org/info/rfc7393>>.
- [RFC7422] Donley, C., Grundemann, C., Sarawat, V., Sundaresan, K., and O. Vautrin, "Deterministic Address Mapping to Reduce Logging in Carrier-Grade NAT Deployments", RFC 7422, DOI 10.17487/RFC7422, December 2014, <<https://www.rfc-editor.org/info/rfc7422>>.
- [RFC7596] Cui, Y., Sun, Q., Boucadair, M., Tsou, T., Lee, Y., and I. Farrer, "Lightweight 4over6: An Extension to the Dual-Stack Lite Architecture", RFC 7596, DOI 10.17487/RFC7596, July 2015, <<https://www.rfc-editor.org/info/rfc7596>>.
- [RFC7597] Troan, O., Ed., Dec, W., Li, X., Bao, C., Matsushima, S., Murakami, T., and T. Taylor, Ed., "Mapping of Address and Port with Encapsulation (MAP-E)", RFC 7597, DOI 10.17487/RFC7597, July 2015, <<https://www.rfc-editor.org/info/rfc7597>>.

- [RFC7599] Li, X., Bao, C., Dec, W., Ed., Troan, O., Matsushima, S., and T. Murakami, "Mapping of Address and Port using Translation (MAP-T)", RFC 7599, DOI 10.17487/RFC7599, July 2015, <<https://www.rfc-editor.org/info/rfc7599>>.
- [RFC7605] Touch, J., "Recommendations on Using Assigned Transport Port Numbers", BCP 165, RFC 7605, DOI 10.17487/RFC7605, August 2015, <<https://www.rfc-editor.org/info/rfc7605>>.
- [RFC7757] Anderson, T. and A. Leiva Popper, "Explicit Address Mappings for Stateless IP/ICMP Translation", RFC 7757, DOI 10.17487/RFC7757, February 2016, <<https://www.rfc-editor.org/info/rfc7757>>.
- [RFC7915] Bao, C., Li, X., Baker, F., Anderson, T., and F. Gont, "IP/ICMP Translation Algorithm", RFC 7915, DOI 10.17487/RFC7915, June 2016, <<https://www.rfc-editor.org/info/rfc7915>>.
- [RFC7950] Bjorklund, M., Ed., "The YANG 1.1 Data Modeling Language", RFC 7950, DOI 10.17487/RFC7950, August 2016, <<https://www.rfc-editor.org/info/rfc7950>>.
- [RFC8114] Boucadair, M., Qin, C., Jacquenet, C., Lee, Y., and Q. Wang, "Delivery of IPv4 Multicast Services to IPv4 Clients over an IPv6 Multicast Network", RFC 8114, DOI 10.17487/RFC8114, March 2017, <<https://www.rfc-editor.org/info/rfc8114>>.
- [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>>.
- [RFC8219] Georgescu, M., Pislaru, L., and G. Lencse, "Benchmarking Methodology for IPv6 Transition Technologies", RFC 8219, DOI 10.17487/RFC8219, August 2017, <<https://www.rfc-editor.org/info/rfc8219>>.
- [RFC8415] Mrugalski, T., Siodelski, M., Volz, B., Yourtchenko, A., Richardson, M., Jiang, S., Lemon, T., and T. Winters, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 8415, DOI 10.17487/RFC8415, November 2018, <<https://www.rfc-editor.org/info/rfc8415>>.

- [RFC8512] Boucadair, M., Ed., Sivakumar, S., Jacquenet, C., Vinapamula, S., and Q. Wu, "A YANG Module for Network Address Translation (NAT) and Network Prefix Translation (NPT)", RFC 8512, DOI 10.17487/RFC8512, January 2019, <<https://www.rfc-editor.org/info/rfc8512>>.
- [RFC8658] Jiang, S., Ed., Fu, Y., Ed., Xie, C., Li, T., and M. Boucadair, Ed., "RADIUS Attributes for Software Mechanisms Based on Address plus Port (A+P)", RFC 8658, DOI 10.17487/RFC8658, November 2019, <<https://www.rfc-editor.org/info/rfc8658>>.
- [RFC8683] Palet Martinez, J., "Additional Deployment Guidelines for NAT64/464XLAT in Operator and Enterprise Networks", RFC 8683, DOI 10.17487/RFC8683, November 2019, <<https://www.rfc-editor.org/info/rfc8683>>.

9.2. Informative References

- [aftr] ISC, "ISC implementation of AFTR", 2022, <<https://www.isc.org/downloads/>>.
- [Azz2021] Al-Azzawi, A. and G. Lencse, "Identification of the Possible Security Issues of the 464XLAT IPv6 Transition Technology", Infocommunications Journal, vol. 13, no. 4, pp. 10-18, DOI: 10.36244/ICJ.2021.4.2, December 2021, <https://www.infocommunications.hu/2021_4_2>.
- [I-D.ietf-v6ops-464xlat-optimization] Martinez, J. P. and A. D'Egidio, "464XLAT/MAT-T Optimization", Work in Progress, Internet-Draft, draft-ietf-v6ops-464xlat-optimization-03, 28 July 2020, <<https://www.ietf.org/archive/id/draft-ietf-v6ops-464xlat-optimization-03.txt>>.
- [I-D.lencse-v6ops-transition-benchmarking] Lencse, G., "Performance Analysis of IPv6 Transition Technologies for IPv4aaS", Work in Progress, Internet-Draft, draft-lencse-v6ops-transition-benchmarking-00, 16 October 2021, <<https://www.ietf.org/archive/id/draft-lencse-v6ops-transition-benchmarking-00.txt>>.
- [I-D.lencse-v6ops-transition-scalability] Lencse, G., "Scalability of IPv6 Transition Technologies for IPv4aaS", Work in Progress, Internet-Draft, draft-lencse-v6ops-transition-scalability-02, 7 March 2022, <<https://www.ietf.org/archive/id/draft-lencse-v6ops-transition-scalability-02.txt>>.

- [jool] NIC.MX, "Open Source SIIT and NAT64 for Linux", 2022, <<http://www.jool.mx>>.
- [LEN2018] Lencse, G. and Y. Kadobayashi, "Methodology for the identification of potential security issues of different IPv6 transition technologies: Threat analysis of DNS64 and stateful NAT64", Computers & Security (Elsevier), vol. 77, no. 1, pp. 397-411, DOI: 10.1016/j.cose.2018.04.012, 1 August 2018, <<http://www.hit.bme.hu/~lencse/publications/ECS-2018-Methodology-revised.pdf>>.
- [LEN2019] Lencse, G. and Y. Kadobayashi, "Comprehensive Survey of IPv6 Transition Technologies: A Subjective Classification for Security Analysis", IEICE Transactions on Communications, vol. E102-B, no.10, pp. 2021-2035., DOI: 10.1587/transcom.2018EBR0002, 1 October 2019, <http://www.hit.bme.hu/~lencse/publications/e102-b_10_2021.pdf>.
- [LEN2020a] Lencse, G., "Benchmarking Stateless NAT64 Implementations with a Standard Tester", Telecommunication Systems, vol. 75, pp. 245-257, DOI: 10.1007/s11235-020-00681-x, 15 June 2020, <<https://link.springer.com/article/10.1007/s11235-020-00681-x>>.
- [LEN2020b] Lencse, G., "Adding RFC 4814 Random Port Feature to Siitperf: Design, Implementation and Performance Estimation", International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems, vol 9, no 3, pp. 18-26, DOI: 10.11601/ijates.v9i3.291, 2020, <<http://ijates.org/index.php/ijates/article/view/291>>.
- [LEN2021] Lencse, G., "Design and Implementation of a Software Tester for Benchmarking Stateless NAT64 Gateways", IEICE Transactions on Communications, DOI: 10.1587/transcom.2019EBN0010, 2021, <https://www.jstage.jst.go.jp/article/transcom/E104.B/2/E104.B_2019EBN0010/_article>.
- [MIY2010] Miyakawa, S., "IPv4 to IPv6 transformation schemes", IEICE Trans. Commun., vol.E93-B, no.5, pp. 1078-1084, DOI:10.1587/transcom.E93.B.10, May 2010, <https://www.jstage.jst.go.jp/article/transcom/E93.B/5/E93.B_5_1078/_article>.

- [REP2014] Repas, S., Hajas, T., and G. Lencse, "Port number consumption of the NAT64 IPv6 transition technology", Proc. 37th Internat. Conf. on Telecommunications and Signal Processing (TSP 2014), Berlin, Germany, DOI: 10.1109/TSP.2015.7296411, July 2014, <<http://www.hit.bme.hu/~lencse/publications/TSP-2014-PC.pdf>>.
- [SIITperf] Lencse, G., "Siitperf: an RFC 8219 compliant SIIT (stateless NAT64) tester", November 2019, <<https://github.com/lencsegabor/siitperf>>.
- [snabb] Igalia, "Snabb implementation of lwAFTR", 2022, <<https://github.com/Igalia/snabb>>.
- [vpp] "VPP Implementations of IPv6-only with IPv4aaS", 2022, <<https://gerrit.fd.io/r/#/admin/projects/>>.

Appendix A. Change Log

A.1. 01 - 02

- * Ian Farrer has joined us as an author.
- * Restructuring: the description of the five IPv4aaS technologies was moved to a separate section.
- * More details and figures were added to the description of the five IPv4aaS technologies.
- * Section titled "High-level Architectures and their Consequences" has been completely rewritten.
- * Several additions/clarification throughout Section titled "Detailed Analysis".
- * Section titled "Performance Analysis" was dropped due to lack of results yet.
- * Word based text ported to XML.
- * Further text cleanups, added text on state sync and load balancing. Additional comments inline that should be considered for future updates.

A.2. 02 - 03

- * The suggestions of Mohamed Boucadair are incorporated.
- * New considerations regarding possible optimizations.

A.3. 03 - 04

- * Section titled "Performance Analysis" was added. It mentions our new benchmarking tool, siitperf, and highlights our plans.
- * Some references were updated or added.

A.4. 04 - 05

- * Some references were updated or added.

A.5. 05 - 06

- * Some references were updated or added.

A.6. 06 - 00-WG Item

- * Stats dated and added for Broadband deployments.
- * Other clarifications and references.
- * New section: IPv4 Pool Size.
- * Typos.

A.7. 00 - 01

To facilitate WGLC, the unfinished parts were moved to two new drafts:

- * New I-D for scale up measurements. (Including the results of iptables.)
- * New I-D for benchmarking measurements. (Only a stub.)

A.8. 01 - 02

Update on the basis of the AD review.

Update of the references.

A.9. 02 - 03

Nits and changes from IESG review.

Updated wrong reference to PCP.

Authors' Addresses

Gabor Lencse
Budapest University of Technology and Economics
Budapest
Magyar tudosok korutja 2.
H-1117
Hungary
Email: lencse@hit.bme.hu

Jordi Palet Martinez
The IPv6 Company
Molino de la Navata, 75
28420 La Navata - Galapagar Madrid
Spain
Email: jordi.palet@theipv6company.com
URI: <http://www.theipv6company.com/>

Lee Howard
Retevia
9940 Main St., Suite 200
Fairfax, Virginia 22031
United States of America
Email: lee@asgard.org

Richard Patterson
Sky UK
1 Brick Lane
London
EQ 6PU
United Kingdom
Email: richard.patterson@sky.uk

Ian Farrer
Deutsche Telekom AG
Landgrabenweg 151
53227 Bonn
Germany

Email: ian.farrer@telekom.de

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 24 February 2022

S. Peng
Z. Li
Huawei Technologies
C. Xie
China Telecom
Z. Qin
China Unicom
G. Mishra
Verizon Inc.
23 August 2021

Processing of the Hop-by-Hop Options Header
draft-peng-v6ops-hbh-06

Abstract

This document describes the processing of the Hop-by-Hop Options Header (HBH) in today's routers in the aspects of standards specification, common implementations, and default operations. This document outlines the reasons why the Hop-by-Hop Options Header is rarely utilized in current networks. In addition, this document describes how the HBH could be used as a powerful mechanism allowing deployment and operations of new services requiring a more optimized way to leverage network resources of an infrastructure. The Hop-by-Hop Options Header is taken into consideration by several network operators as a valuable container for carrying the information facilitating the introduction of new services. The processing requirements of the HBH and the migration strategies are also suggested.

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 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

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 24 February 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Modern Router Architecture	4
3. Specification of RFC 8200	7
4. Common Implementations	8
4.1. Historical Reasons	9
4.2. Consequences	9
5. Operators' Typical Processing	9
6. New Services	10
7. Requirements	10
8. Migration Strategies	11
9. Security Considerations	12
10. IANA Considerations	12
11. Acknowledgements	12
12. References	12
12.1. Normative References	12
12.2. Informative References	13
Authors' Addresses	14

1. Introduction

Due to historical reasons, such as incapable ASICs, limited IPv6 deployments, and few service requirements, the most common Hop-by-Hop Options header (HBH) processing implementation is that the node sends the IPv6 packets with the Hop-by-Hop Options header to the control plane of the node. The option type of each option carried within the Hop-by-Hop Options header will not even be examined before the packet is sent to the control plane. Very often, such processing behavior is the default configuration or, even worse, is the only behavior of the ipv6 implementation of the node.

Such default processing behavior of the Hop-by-Hop Options header could result in various unpleasant effects such as a risk of Denial of Service (DoS) attack on the router control plane and inconsistent packet drops due to rate limiting on the interface between the router control plane and forwarding plane, which will impact the normal end-to-end IP forwarding of the network services.

This actually introduced a circular problem:

-> An implementation problem caused HBH to become a DoS vector.

-> Because HBH is a DoS vector, network operators deployed ACLs that discard packets containing HBH.

-> Because network operators deployed ACLs that discard packets containing HBH, network designers stopped defining new HBH Options.

-> Because network designers stopped defining new HBH Options, the community was not motivated to fix the implementation problem that cause HBH to become a DoS vector.

The purpose of this draft is to break the cycle described above, fixing the problem that caused HBH not actually being utilized in operators' networks so to allow a better leverage of the HBH capability.

Driven by the wide deployments of IPv6 and ever-emerging new services, the Hop-by-Hop Options Header is taken as a valuable container for carrying the information to facilitate these new services.

This document suggests the desired processing behavior and the migration strategies towards it.

2. Modern Router Architecture

Modern router architecture design maintains a strict separation of the router control plane and its forwarding plane [RFC6192], as shown in Figure 1. Either the control plane or the forwarding plane is composed of both software and hardware, but each plane is responsible for different functions.

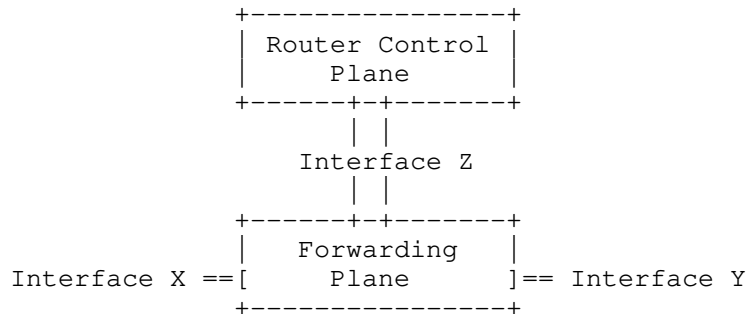


Figure 1. Modern Router Architecture

The router control plane supports routing and management functions, handling packets destined to the device as well as building and sending packets originated locally on the device, and also drives the programming of the forwarding plane. The router control plane is generally realized in software on general-purpose processors, and its hardware is usually not optimized for high-speed packet handling. Because of the wide range of functionality, it is more susceptible to security vulnerabilities and a more likely a target for a DoS attack.

The forwarding plane is typically responsible for receiving a packet on an incoming interface, performing a lookup to identify the packet's next hop and determine the outgoing interface towards the destination, and forwarding the packet out through the appropriate outgoing interface. Typically, forwarding plane functionality is realized in high-performance Application Specific Integrated Circuits (ASICs) or Network Processors (NPs) that are capable of handling very high packet rates.

The router control plane interfaces with its forwarding plane through the Interface Z, as shown in the Figure 1, and the forwarding plane connects to other network devices via Interfaces such as X and Y. Since the router control plane is vulnerable to the DoS attack, usually a traffic filtering mechanism is implemented on Interface Z in order to block unwanted traffic. In order to protect the router control plane, a rate-limiting mechanism is always implemented on this interface. However, such rate limiting mechanism will always cause inconsistent packet drops, which will impact the normal IP forwarding.

Semiconductor chip technology has advanced significantly in the last decade, and as such the widely used network processing and forwarding process can now not only forward packets at line speed, but also easily support other feature processing such as QoS for DiffServ/MPLS, Access Control List (ACL), Firewall, and Deep Packet Inspection (DPI).

A Network Processing Unit (NPU) is a non-ASIC based Integrated Circuit (IC) that is programmable through software. It performs all packet header operations between the physical layer interface and the switching fabric such as packet parsing and forwarding, modification, and forwarding. Many equipment vendors implement these functions in fixed function ASICs rather than using "off-the-shelf" NPUs, because of proprietary algorithms.

Classification Co-processor is a specialized processor that can be used to lighten the processing load on an NPU by handling the parsing and classification of incoming packets such as IPv6 extended header HBH options processing. This advancement enables network processors to do the general process to handle simple control messages for traffic management, such as signaling for hardware programming, congestion state report, OAM, etc. Industry trend is for intelligent multi-core CPU hardware using modern NPUs for forwarding packets at line rate while still being able to perform other complex tasks such as HBH forwarding options processing without having to punt to the control plane.

Many of the packet-processing devices employed in modern switch and router designs are fixed-function ASICs to handle proprietary functions. While these devices can be very efficient for the set of functions they are designed for, they can be very inflexible. There is a tradeoff of price, performance and flexibility when vendors make a choice to use a fixed function ASIC as opposed to NPU. Due to the inflexibility of the fixed function ASIC, tasks that require additional processing such as IPv6 HBH header processing must be punted to the control plane. This problem is still a challenge today and is the reason why operators to protect against control plane DOS

attack vector must drop or ignore HBH options. As industry shifts to Merchant Silicon based NPU evolution from fixed function ASIC, the gap will continue to close increasing the viability ubiquitous HBH use cases due to now processing in the forwarding plane.

Most modern routers maintain a strict separation between forwarding plane and control plane hardware. Forwarding plane bandwidth and resources are plentiful, while control plane bandwidth and resources are constrained. In order to protect scarce control plane resources, routers enforce policies that restrict access from the forwarding plane to the control plane. Effective policies address packets containing the HBH Options Extension header, because HBH control options require access from the forwarding plane to the control plane. Many network operators perceive HBH Options to be a breach of the separation between the forwarding and control planes. In this case HBH control options would be required to be punted to control plane by fixed function ASICs as well as NPUs.

The maximum length of an HBH Options header is 2,048 bytes. A source node can encode hundreds of options in 2,048 bytes. With today's technology it would be cost prohibitive to be able to process hundreds of options with either NPU or proprietary fixed function ASIC.

While [RFC8200] required that all nodes must examine and process the Hop-by-Hop Options header, it is now expected that nodes along a packet's delivery path only examine and process the Hop-by-Hop Options header if explicitly configured to do so. This can be beneficial in cases where transit nodes are legacy hardware and the destination endpoint PE is newer NPU based hardware that can process HBH in the forwarding plane.

IPv6 Extended Header limitations that need to be addressed to make HBH processing more efficient and viable in the forwarding plane:

[RFC8504] defines the IPv6 node requirements and how to protect a node from excessive header chain and excessive header options with various limitations that can be defined on a node. [RFC8883] defines ICMPv6 Errors for discarding packets due to processing limits. Per [RFC8200] HBH options must be processed serially. However, an implementation of options processing can be made to be done with more parallelism in serial processing grouping of similar options to be processed in parallel.

The IPv6 standard does not currently limit the header chain length or number of options that can be encoded.

Each Option is encoded in a TLV and so processing of a long list of TLVs is expensive. Zero data length encoded options TLVs are a valid option. A DOS vector could be easily generated by encoding 1000 HBH options (Zero data length) in a standard 1500 MTU packet. So now imagine if you have a Christmas tree long header chain to parse each with many options.

3. Specification of RFC 8200

[RFC8200] defines several IPv6 extension header types, including the Hop-by-Hop (HBH) Options header. As specified in [RFC8200], the Hop-by-Hop (HBH) Options header is used to carry optional information that will be examined and processed by every node along a packet's delivery path, and it is identified by a Next Header value of zero in the IPv6 header.

The Hop-by-Hop (HBH) Options header contains the following fields:

-- Next Header: 8-bit selector, identifies the type of header immediately following the Hop-by-Hop Options header.

-- Hdr Ext Len: 8-bit unsigned integer, the length of the Hop-by-Hop Options header in 8-octet units, not including the first 8 octets.

-- Options: Variable-length field, of length such that the complete Hop-by-Hop Options header is an integer multiple of 8 octets long.

The Hop-by-Hop (HBH) Options header carries a variable number of "options" that are encoded in the format of type-length-value (TLV).

The highest-order two bits (i.e., the ACT bits) of the Option Type specify the action that must be taken if the processing IPv6 node does not recognize the Option Type. The third-highest-order bit (i.e., the CHG bit) of the Option Type specifies whether or not the Option Data of that option can change en route to the packet's final destination.

While [RFC2460] required that all nodes must examine and process the Hop-by-Hop Options header, with [RFC8200] it is expected that nodes along a packet's delivery path only examine and process the Hop-by-Hop Options header if explicitly configured to do so. It means that the HBH processing behavior in a node depends on its configuration.

However, in the current [RFC8200], there is no explicit specification of the possible configurations. Therefore, the nodes may be configured to ignore the Hop-by-Hop Options header, drop packets containing a Hop-by-Hop Options header, or assign packets containing a Hop-by-Hop Options header to the control plane [RFC8200]. Because of these likely uncertain processing behaviors, new hop-by-hop options are not recommended.

4. Common Implementations

In the current common implementations, once an IPv6 packet, with its Next Header field set to 0, arrives at a node, it will be directly sent to the control plane of the node. With such implementations, the value of the Next Header field in the IPv6 header is the only trigger for the default processing behavior. The option type of each option carried within the Hop-by-Hop Options header will not even be examined before the packet is sent to the control plane.

Very often, such processing behavior is the default configuration on the node, which is embedded in the implementation and cannot be changed or reconfigured.

Another critical component of IPv6 HBH processing, in some cases overlooked, is the operator core network which can be designed to use the global Internet routing table for internet traffic and in other cases use an overlay MPLS VPN to carry Internet traffic.

In the global Internet routing table scenario where only an underlay global routing table exists, and no VPN overlay carrying customer Internet traffic, the IPv6 HBH options can be used as a DOS attack vector for both the operator nodes, adjacent inter-as peer nodes as well as customer nodes along a path.

In a case where the Internet routing table is carried in a MPLS VPN overlay payload, the HBH options header does not impact the operator underlay framework and only impacts the VPN overlay payload and thus the operator underlay topmost label global table routing FEC LSP instantiation is not impacted as the operator underlay is within the operators closed domain.

However, HBH options DOS attack vector in the VPN overlay can still impact the customer CE destination end nodes as well as other adjacent inter-as operators that only use underlay global Internet routing table. In an operator closed domain where MPLS VPN overlay is utilized to carry internet traffic, the operator has full control of the underlay and IPv6 Extended header chain length as well as the number of HBH options encoded.

In the global routing table scenario for Internet traffic there is no way to control the IPv6 Extended header chain length as well as the number of HBH options encoded.

4.1. Historical Reasons

When IPv6 was first implemented on high-speed routers, HBH options were not yet well-understood and ASICs were not as capable as they are today. So, early IPv6 implementations dispatched all packets that contain HBH options to their control plane.

4.2. Consequences

Such implementation introduces a risk of a DoS attack on the control plane of the node, and a large flow of IPv6 packets could congest the control plane, causing other critical functions (including routing and network management) that are executed on the control plane to fail. Rate limiting mechanisms will cause inconsistent packet drops and impact the normal end-to-end IP forwarding of the network services.

5. Operators' Typical Processing

To mitigate this DoS vulnerability, many operators deployed Access Control Lists (ACLs) that discard all packets containing HBH Options.

[RFC6564] shows the Reports from the field indicating that some IP routers deployed within the global Internet are configured either to ignore or to drop packets having a hop-by-hop header. As stated in [RFC7872], many network operators perceive HBH Options to be a breach of the separation between the forwarding and control planes. Therefore, several network operators configured their nodes so as to discard all packets containing the HBH Options Extension Header, while others configured nodes to forward the packet but to ignore the HBH Options. [RFC7045] also states that hop-by-hop options are not handled by many high-speed routers or are processed only on a control plane.

Due to such behaviors observed and described in these specifications, new hop-by-hop options are not recommended in [RFC8200] hence the usability of HBH options is severely limited.

6. New Services

As IPv6 is being rapidly and widely deployed worldwide, more and more applications and network services are migrating to or directly adopting IPv6. More and more new services that require HBH are emerging and the HBH Options header is going to be utilized by the new services in various scenarios.

In-situ OAM (IOAM) with IPv6 encapsulation [I-D.ietf-ippm-ioam-ipv6-options] is one of the examples. IOAM in IPv6 is used to enhance diagnostics of IPv6 networks and complements other mechanisms, such as the IPv6 Performance and Diagnostic Metrics Destination Option described in [RFC8250]. The IOAM data fields are encapsulated in "option data" fields of the Hop-by-Hop Options header if Pre-allocated Tracing Option, Incremental Tracing Option, or Proof of Transit Option are carried [I-D.ietf-ippm-ioam-data], that is, the IOAM performs per hop.

Alternate Marking Method can be used as the passive performance measurement tool in an IPv6 domain. The AltMark Option is defined as a new IPv6 extension header option to encode alternate marking technique and Hop-by-Hop Options Header is considered [I-D.ietf-6man-ipv6-alt-mark].

The Minimum Path MTU Hop-by-Hop Option is defined in [I-D.ietf-6man-mtu-option] to record the minimum Path MTU along the forward path between a source host to a destination host. This Hop-by-Hop option is intended to be used in environments like Data Centers and on paths between Data Centers as well as other environments including the general Internet. It provides a useful tool for allowing to better take advantage of paths able to support a large Path MTU.

As more services start utilizing the HBH Options header, more packets containing HBH Options are going to be injected into the networks. According to the current common configuration in most network deployments, all the packets of the new services are going to be sent to the control plane of the nodes, with the possible consequence of causing a DoS on the control plane. The packets will be dropped and the normal IP forwarding may be severely impacted. The deployment of new network services involving multi-vendor interoperability will become impossible.

7. Requirements

- * The HBH options header MUST NOT become a possible DDoS Vector.

- * HBH options SHOULD be designed so that they don't reduce the probability of packet delivery. For example, an intermediate node may discard a packet because it contains more HBH options than the node can process.
- * HBH processing MUST be efficient. That is, it MUST be possible to produce implementations that perform well at a reasonable cost.
- * The Router Alert Option MUST NOT impact the processing of other HBH options that should be processed more quickly.
- * HBH Options MAY influence how a packet is forwarded. However, with the exception of the Router Alert Option, an HBH Option MUST NOT cause control plane state to be created, modified or destroyed on the processing node. As per [RFC6398], protocol developers SHOULD avoid future use of the Router Alert Option.
- * More requirements are to be added.

8. Migration Strategies

In order to achieve the desired processing behavior of the HBH options header and facilitate the ever-emerging new services to be deployed in operators' networks across multiple vendors' devices, the migration can happen in three parts as described below:

1. The source of the HBH options header encapsulation.

The information to be carried in the HBH options header needs to be first categorized and encapsulated into either control options or forwarding options, and then encapsulated in different packets.

2. The nodes within the network.

The nodes within the network are updated to the proposed behavior introduced in the previous section.

3. The edge nodes of the network.

The edge nodes should check whether the packet contains an HBH header with control or forwarding option. Packets with a control option may still be filtered and dropped while packets with forwarding option SHOULD be allowed by the ACL.

If it is certain that there is no harm that can be introduced by the HBH control options to the nodes and the services, they can also be allowed.

Note: During the migration stage, the nodes that are not yet updated will stay with their existing configurations.

9. Security Considerations

The same as the Security Considerations apply as in [RFC8200] for the part related with the HBH Options header.

10. IANA Considerations

This document does not include an IANA request.

11. Acknowledgements

The authors would like to acknowledge Ron Bonica, Fred Baker, Bob Hinden, Stefano Previdi, and Donald Eastlake for their valuable review and comments.

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, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <<https://www.rfc-editor.org/info/rfc2460>>.
- [RFC6192] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, DOI 10.17487/RFC6192, March 2011, <<https://www.rfc-editor.org/info/rfc6192>>.
- [RFC6398] Le Faucheur, F., Ed., "IP Router Alert Considerations and Usage", BCP 168, RFC 6398, DOI 10.17487/RFC6398, October 2011, <<https://www.rfc-editor.org/info/rfc6398>>.
- [RFC7045] Carpenter, B. and S. Jiang, "Transmission and Processing of IPv6 Extension Headers", RFC 7045, DOI 10.17487/RFC7045, December 2013, <<https://www.rfc-editor.org/info/rfc7045>>.

- [RFC7872] Gont, F., Linkova, J., Chown, T., and W. Liu, "Observations on the Dropping of Packets with IPv6 Extension Headers in the Real World", RFC 7872, DOI 10.17487/RFC7872, June 2016, <<https://www.rfc-editor.org/info/rfc7872>>.
- [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>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

12.2. Informative References

- [I-D.ietf-6man-ipv6-alt-mark]
Fioccola, G., Zhou, T., Cociglio, M., Qin, F., and R. Pang, "IPv6 Application of the Alternate Marking Method", Work in Progress, Internet-Draft, draft-ietf-6man-ipv6-alt-mark-08, 26 July 2021, <<https://www.ietf.org/archive/id/draft-ietf-6man-ipv6-alt-mark-08.txt>>.
- [I-D.ietf-6man-mtu-option]
Hinden, R. M. and G. Fairhurst, "IPv6 Minimum Path MTU Hop-by-Hop Option", Work in Progress, Internet-Draft, draft-ietf-6man-mtu-option-06, 7 August 2021, <<https://www.ietf.org/archive/id/draft-ietf-6man-mtu-option-06.txt>>.
- [I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-data-14, 24 June 2021, <<https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-data-14.txt>>.
- [I-D.ietf-ippm-ioam-ipv6-options]
Bhandari, S. and F. Brockners, "In-situ OAM IPv6 Options", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-ipv6-options-06, 31 July 2021, <<https://www.ietf.org/archive/id/draft-ietf-ippm-ioam-ipv6-options-06.txt>>.

- [RFC2711] Partridge, C. and A. Jackson, "IPv6 Router Alert Option", RFC 2711, DOI 10.17487/RFC2711, October 1999, <<https://www.rfc-editor.org/info/rfc2711>>.
- [RFC8250] Elkins, N., Hamilton, R., and M. Ackermann, "IPv6 Performance and Diagnostic Metrics (PDM) Destination Option", RFC 8250, DOI 10.17487/RFC8250, September 2017, <<https://www.rfc-editor.org/info/rfc8250>>.
- [RFC8504] Chown, T., Loughney, J., and T. Winters, "IPv6 Node Requirements", BCP 220, RFC 8504, DOI 10.17487/RFC8504, January 2019, <<https://www.rfc-editor.org/info/rfc8504>>.
- [RFC8883] Herbert, T., "ICMPv6 Errors for Discarding Packets Due to Processing Limits", RFC 8883, DOI 10.17487/RFC8883, September 2020, <<https://www.rfc-editor.org/info/rfc8883>>.

Authors' Addresses

Shuping Peng
Huawei Technologies
Beijing
China

Email: pengshuping@huawei.com

Zhenbin Li
Huawei Technologies
Beijing
China

Email: lizhenbin@huawei.com

Chongfeng Xie
China Telecom
China

Email: xiechf@chinatelecom.cn

Zhuangzhuang Qin
China Unicom
Beijing
China

Email: qinzhuangzhuang@chinaunicom.cn

Gyan Mishra
Verizon Inc.
United States of America

Email: gyan.s.mishra@verizon.com

IPv6 Operations (v6ops) Working Group
Internet Draft
Intended status: Informational
Expires: March 2022

E. Vasilenko
X. Xiao
Huawei Technologies
D. Khaustov
Rostelecom
September 17, 2021

IPv6 Oversized Packets Analysis
draft-vasilenko-v6ops-ipv6-oversized-analysis-01

Abstract

The IETF has some initiatives relying on IPv6 Extension Headers added in transit: SRv6, iOAM. Additionally, some recent developments are overlays (SRv6, VxLAN, NVO3, L2TPv3, and LISP). It could create oversized packets that need to be dealt with. This document analyzes available standards for the resolution of oversized packet drops.

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 February 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Terminology and pre-requisite.....	2
2. Problem statement.....	3
3. Solutions.....	5
3.1. Provision links with big enough MTU.....	6
3.2. Frugal usage of Extension Headers.....	7
3.3. Fragmentation and reassembly at the tunnel ends.....	9
3.4. PMTUD by original packet source.....	12
3.5. Packetization Layer MTU Discovery.....	14
4. Conclusion.....	15
5. Security Considerations.....	15
6. IANA Considerations.....	15
7. References.....	16
7.1. Normative References.....	16
7.2. Informative References.....	18
8. Acknowledgments.....	19

1. Terminology and pre-requisite

We do assume good knowledge or frequent references to [PMTUD] and [IPv6 Tunneling]. Terminology is inherited from [PMTUD].

Link MTU - the maximum transmission unit, i.e., maximum packet size in octets that can be conveyed over a link.

Path MTU (PMTU) - the minimum link MTU of all links in a path between a source node and a destination node.

Path MTU Discovery (PMTUD) - the process by which a node learns the PMTU of a path.

EMTU_S - Effective MTU for sending; used by upper-layer protocols to limit the size of IP packets they queue for sending.

EMTU_R - Effective MTU for receiving; the largest packet that can be reassembled at the receiver.

Packetization Layer - the layer of the network stack that segments data into packets.

PLPMTUD - Packetization Layer Path MTU Discovery, the method of detecting path MTU at packetization layer, which is an extension of classical PMTU Discovery.

PTB (Packet Too Big) message - an ICMPv6 message reporting that an IPv6 packet is too large to forward through some link.

MSS - the TCP Maximum Segment Size, the maximum payload size available to the TCP layer. This is typically the Path MTU minus the size of the IP and TCP headers.

2. Problem statement

IPv6 is strict regarding fragmentation - it must NOT be done in transit (section 4.5 of [IPv6]).

IPv6 sees rapid developments in recent years. A lot of additional functionality has been added primarily by adding options to Extension Headers and/or using overlay encapsulation. All of the above expand the packet size. This could lead to oversized packets that would be dropped on some links.

Massive parallelism in traffic delivery is the additional challenge developed in the last 10 years: ECMP on one hop could reach 16 (or even more), which creates the end-to-end possibility for 64k paths on just 5 hops (example from big production network). Different paths could have a different set of Extension Headers and different PMTU as a result. PMTU is effectively becoming dynamic: we could never know how many additional headers would be added at a particular time to the particular packet on the particular path.

The old classical PMTUD problems are still with us: filtered ICMPv6 messages, drops related to Extension Headers before next hop MTU has been evaluated (no Packet Too Big message sent).

Standards have two important numbers that we would need for our discussion:

- o [IPv6] chapter 5 requires that every link should have the MTU of 1280 octets or greater ($2^{10}+2^8$ - it probably explains the choice of this size)

- o [IPv6] requests minimum EMTU_R (reassembly buffer) in 1500 octets. An upper-layer protocol or application that depends on IPv6 fragmentation to send packets larger than the MTU of a path should not send packets larger than 1500 octets unless it has the assurance that the destination is capable of reassembling packets of that larger size

The reassembly buffer is much above 1500B for the majority of desktop and server OSes. Windows 10 has "Reassemblylimit" almost 64MB (you could look by "netsh interface ipv6 show global"). Different flavors of Linux have "ipfrag_high_thresh" between 256KB and 4MB (you could look by "more /proc/sys/net/ipv4/ipfrag_high_thresh"). iOS has "maxReceiveIPv6BufferSize" 64Kb.

The reassembly is not so good for embedded OSes. From the four primary OSes for IoT (Contiki, FreeRTOS, Mbed OS, MicroPython), only Mbed OS has the capability for 5 fragments by default, and it is possible to activate reassembly on Contiki. In all cases, the buffer is just a few packets of 1280B or 1500B. IoT devices may not be capable to reassemble the packet that the server in the cloud would send to it. Hence, ICMP PTB is still very important for some OSes.

There is only one solution by [IPv6] architecture for the PMTU problem - decrease packet size on the original source. It is workable up to the minimum limit for IPv6 packets (1280B). The typical transit link had MTU not much bigger than 1500B for a long time, only the space for a few additional MPLS labels was reserved. 220B left could be considered as guaranteed for additional functionality in Extension and Encapsulation headers. It could be enough for the next decade if we would make some precautions - see discussion below.

[Huston-2016] and [Huston-2021] did an investigation on a different topic (fragmentation), but he has good statistics related to MTU drops up to 1500B that did show a 5% drop for MTU as small as 1455B. Additionally, [Huston-2016] has found the big drop spike (69% from all drops!) at 1480B, 20B less is presumable for IPv6 encapsulation into IPv4. [Huston-2021] has shown twice bigger fragmentation drop for bigger packets with the peak at 1408 octets, especially for Asia. As you can see - 1500B is not always available now, the reason is not well understood. Hence, we do not have 220B for additional headers in all situations. We could be reasonably optimistic that such a type of tunneling would disappear in the long term. Later, we would stick to an optimistic assumption that 220B is available in most situations. It is still possible to have the more pessimistic estimation (200B? 175B?) looking to Huston's data.

The hungriest protocol known is SRv6 that could add 40B of IPv6 underlay tunnel header (called "outer IP header" in [SRH]), 16B of SRH header itself, and additionally up to 10 IPv6 addresses in the SID stack (potentially even more). It is already 216B - very close to 220B optimistic limit. It makes the introduction of any additional functionality quite challenging without rigorous expansion of all links to bigger MTU.

Initial SRv6 implementations that trespassed safe limit in 220B are the reason for recent activities in MTU problem research. We see many recent efforts to improve Path MTU Discovery (which would be mentioned in the document) - let us find the rationale behind it.

3. Solutions

Let's consider first the reassembly buffer problem as the simpler one.

Minimal buffer for packet reassembly (1500B) is potentially possible to increase in new standard updates, but then would be the problem with the transition, because this limitation would be already programmed into billions of IoT hosts - it would need big time to be sure that we do not have old implementation anymore.

There is no good solution for the problem of bloating headers above 220B for hosts. We need to keep headers below the 220B limit for embedded OSes. Fortunately, we are far from this problem yet - very limited additional functionality is implemented directly on the hosts (like [PMTU by HbH] or APN6). This problem should be looked at again in some number of years, it may be that in the future we would have to increase default EMTU_R on all hosts to give the possibility for new functionality.

Let's now return to our primary problem of not enough PMTU.

There is a low probability that the Internet community would agree to decrease the minimal IPv6 packet size (1280B). Hence, the oversized problem could not be resolved in that direction.

It is possible to partially alleviate the MTU problem in some network zones where all transit nodes have big enough MTU. Transit nodes should delete extension headers before packets would leave "high MTU network zone". The leakage of a big header to a host could overflow EMTU_R buffer. The majority of RFCs recommend carriers delete additional headers before forwarding traffic to the client - this practice should be strictly followed.

The SPRING working group is actively developing a compressed version of SRv6 that should leave space for other functionality, even on current transit routers that sometimes do not support much above 1500B.

All solutions for packet drop avoidance as a result of oversized packets could be classified into 4 classes. They are examined one by one.

3.1. Provision links with big enough MTU

MTU supported by the host's links is typically 1500 Bytes. Backbone link's MTU could be up to 9000 Bytes on modern hardware. PMTUD is not needed in an ideal world.

Reality is not that good:

- o Some old devices still support just a few additional MPLS labels above 1500B on Ethernet. It was historically a problem to cross 1536B because the IEEE specification for 802.3 assumes that a bigger number in the Length field means the Type of the payload.
- o We could have middleboxes that would not support MTU much bigger than 1500B MTU for a long time.
- o Ethernet is very mature now in relation to big MTU support, but that could be a challenge for other link-layer technologies (for example WiFi, satellite links, radio links, etc.).
- o Packet Links could be rented from a third party - no possibility to change the MTU.
- o Big MTU may negatively influence buffer size - see below.
- o The majority of vendors set the default MTU to 1500B (with variations on what is counted inside MTU). It is time-consuming to change the MTU on the production network.
- o Some hosts (especially for storage traffic in Data Centers) could use 2500B or 9000B MTU that challenges the possibility of having always bigger MTU in the backbone.

Cost-optimized equipment architecture (especially used for switches, but applicable for many routers as well) may not split packets in the buffer memory. So small packet would occupy a bigger buffer space reserved for the packet with maximum MTU. This limitation effectively decreases the potential number of packets that could be buffered. Most of the host packets are still limited to 1500B size. MTU 9000B would just lead to wasting buffer memory with an efficiency of 1/6. The average packet size is twice smaller, hence in the worst-case buffer efficiency could be up to 1/12. Buffer memory is about 30% of the router cost. It is not acceptable to

increase buffer memory cost 12 times. Hence, in many cases, it does not make sense to increase MTU to the maximum supported by the switch or router. One should always check with the vendor the impact of using a big MTU on buffering for the particular product. MTU should be increased to the number that is bigger than the maximum MTU expected from hosts + the size of all possible network overhead + underlay IPv6 header (if present). 9000B MTU makes sense in DC, cross-DC environment, or for platforms that fragment packets for smaller sells in the memory.

[MTU issues in Tunneling] section 3.3 discusses the opposite solution: decrease MTU on links to hosts to be sure that a host would always generate small enough MTU for the backbone. This solution was possible for small tunnel overhead, but now we are talking about the situation when 220B margin is not enough.

[L3VPN] and [EVPN] do attach an additional label and could create oversized packets. Still, the MPLS header cannot point to the original MPLS router that has an attached service label. Additionally, a VPN IP packet could use private address space or no IP address at all (for EVPN). It blocks the possibility to properly organize the PMTUD process. Hence, [L3VPN] and [EVPN] have been developed under the assumption that all MTUs on the path would be expanded for at least 8 bytes that are needed for services over the MPLS data plane.

We have recent [Generic Delivery Functions] that may permit fragmentation for MPLS services, but it is a personal draft yet.

[Pseudowire Fragmentation] is the rear case when fragmentation is available over MPLS for one type of service.

[VxLAN] section 4.3 also uses the approach: "it is RECOMMENDED that the MTUs across the physical network infrastructure be set to a value that accommodates the larger frame size due to the encapsulation".

Packet drop statistics and big activity in IETF prove that the PMTUD problem persists.

"Raise MTU on transit" is the best solution, if it is available.

3.2. Frugal usage of Extension Headers

Some new functionality (especially source routing with a big SID stack) could decrease headers size without a big loss of functionality (for example, use loose node appointment in SID

stack). Some functionality (like iFIT or iOAM) could be completely omitted in the situation that would lead to packet drop. It is effectively "the tradeoff of functionality to PMTU control".

The important point here is that the transit node attaching an additional header should be aware of all MTUs along the assumed packet path to predict how big MTU is still acceptable.

[PMTUD] is readily available for tunneling interfaces - tunnel source should be aware of PMTU of the tunnel (by PTB feedback messages). But we have cases when it is not enough:

- o SDN controller (or management system in general) could assist in provisioning of extension headers (including SFC, iOAM, BIER) and encapsulation headers (SRv6, VxLAN) - should be the way to report MTUs to Controller.
- o ICMPv6 PTB would be directed to the transit control plane only in the case of problems inside the tunnel. PTB messages from outside of the tunnel would be directed to the source node. It is difficult to snoop PTB on transit nodes.

Hence, we see many initiatives to collect and manage MTU by many popular protocols for routing and traffic engineering: [PMTU by ISIS], [PMTU by BGP-LS], [PMTU by PCEP], [PMTU by SR-Policy].

Moreover, these protocol extensions would become even more useful in the future when it would not be possible to squeeze all extension headers into 220B anymore. Frugal attachment of new headers on transit nodes would increase the need for awareness of PMTU - it should stimulate MTU collection by all other popular protocols (OSPF, normal BGP on peering borders).

This approach has a fundamental problem: full knowledge about all MTUs in the domain could not help to estimate the real path for a packet, because of massive ECMP used by many networks (at least by all Carriers). Non-routing protocols do not have a proper engine to estimate traffic paths and predict PMTU as well. And even more, if L2 ECMP is used or some links are rented from another carrier it will again be impossible to predict the exact path and the PMTU.

The second problem of this approach could be classified as "chicken and egg". We already have a much better solution for MTU drop - increase MTU (see the previous section). We are looking for other solutions only because upgrading equipment (to better MTU) is not possible for some reason. But new protocols introduction would also demand equipment upgrade and thus making frugal headers less

valuable. However, an upgrade for the control plane should be cheaper than an upgrade for the data plane, if the vendor would support such an approach.

Hence, the solution discussed in this section has only limited applicability.

3.3. Fragmentation and reassembly at the tunnel ends

The tunnel source behaves like a host with respect to the tunnel header. It is possible to properly adjust PMTU for the tunnel by [PMTUD], so it is potentially possible to fragment all packets bigger than PMTU.

[IP Encapsulation] is the earliest standard for IP-in-IP encapsulation. Section 5.1 discusses that it is possible to fragment IP packets before tunnel encapsulation, so there is no need to reassemble packets on the other tunnel end - reassembly could happen on the destination host. It does not have additional cost implications on tunnel ends. This approach did work for IPv4 in the case of the "don't fragment" bit cleared. It fully contradicts IPv6 architecture that does not permit to fragment packets on transit - no standard has risked proposing such a solution for IPv6.

Some standards do propose IPv6 fragmentation (primarily for packets 1280B and below), but fragmentation is recommended after encapsulation. It would lead to packet reassembly on the other tunnel end to hide (from destination host) the fact of transit fragmentation. It does minimize IPv6 architecture disruption.

Many standards discussed below ([MPLS Encapsulation], [L2TPv3], [VxLAN], [NVO3]) forgot to mention that packets 1280 and below should be fragmented. This inaccuracy did not create any problem in production networks because we typically have 220B for all headers - it is big enough for many tunnels nested into each other. The situation could change in the next years because of Extension Headers expansion by different functions. It could create pressure to return to many mature standards and clarify the situation: what to do when the 1280B packet could not go through the tunnel.

The Fragmentation has a few issues that make it not popular:

- o Fragmentation could double buffer requirements (we assume split only in 2 fragments). We could ignore small additional buffer requirements for packets that may be lost and need to wait some time before reassembly, the Internet is not productive anyway

after a few percentages of packet drops. The buffer memory is about 30% of the router cost. A 30% cost increase would not be accepted by the majority of owners. Albeit, some middleboxes already have enough buffer memory that could be reused for packet reassembly.

- o In general, IPv6 architecture does not approve fragmentation in transit in all standards (except the recent draft [IP Tunnels] - see below). [PMTUD] section 5.1: "packetization layers are encouraged to avoid sending messages that will require fragmentation".
We would discuss in this section some situations when tunnel fragmentation is inevitable.
- o [Fragile Fragmentation] has a good collection of all problems related to fragmentation (additionally to the above: breaks ECMP, stateful processing, policy routing, and has many security attack vectors). [Fragile Fragmentation] strongly recommends avoiding fragmentation, but not deprecate yet.

The primary RFC for tunneling is [IPv6 Tunneling] - it is the oldest standard that was later reused by many other standards (including the latest SRH). It permits fragmentation only for the case when the original packet is already minimal (1280B or less) - see section 7.1. It mandates dropping the packet and signaling ICMPv6 PTB to the source (request to decrease the PMTU size at the source) for all other cases.

[MPLS Encapsulation] Section 5.1 has the name: "Preventing Fragmentation and Reassembly". It does stress again: "IPv6 intermediate nodes do not perform fragmentation in any event".

[L2TPv3] section 4.1.4 has a similar comment: "Note that IPv6 does not support "in-flight" fragmentation of data packets".

[VxLAN] section 4.3 is strict: "VTEPs MUST NOT fragment VXLAN packets."

[NVO3] section 4.4.4 is strict too: "It is strongly RECOMMENDED that Path MTU Discovery ([PMTUD]) be used to prevent or minimize fragmentation."

[IPv6 GRE] section 3.3 does recommend fragmentation only for packets that are less than 1280B.

The most recent draft for all types of tunnels is [IP Tunnels]. It is already referenced by many IETF documents. It is complicated to cover all use cases (any IP over any IP in any situation), but the

net result is: much bigger part of the traffic proposed to be fragmented into the tunnel. Section 3.3: "The path between ingress and egress interfaces has a path MTU, but the endpoints can exchange messages as large as can be reassembled at the destination (egress interface), i.e., the EMTU_R of the egress interface".

A short explanation of proposed functionality: original host would try to transmit biggest flows (by volume) on maximum PMTU, that tunnel source would not try to correct by PTB messages up to 1500B. Hence, the tunnel source would not have any option except to fragment. The principal problem here is the absence of PTB messages for the packet size between PMTU and statically appointed EMTU_R. Let's see how it has been formulated in more detail.

[IP Tunnels] introduces a new variable "Tunnel MTU" that should not change as a result of PMTUD. The procedure to change "Tunnel MTU" is out of the draft discussion - it is pushed to specifications of particular tunnels in the last paragraph of section 4.2.2. Moreover, it is even assumed that PLPMTUD could be used on the router for "Tunnel MTU" discovery because this parameter is considered as an above network layer (like transport layer on the host). Separate section 4.2.3 is dedicated to the explanation that the newly introduced "Tunnel MTU" cannot be adjusted dynamically. There is a recommendation for the default "Tunnel MTU": typical host EMTU_R (1500B) minus tunnel outer headers overhead. The good question could be: if it is so difficult to manage "Tunnel MTU" dynamically, then why this variable was introduced?

The MTU of the tunnel is renamed into MAP (maximum atomic packet), MAP should be corrected by PMTUD feedback from inside the tunnel. Section 4.2.2 states that everything up to "Tunnel MTU" should be accepted to the tunnel, one long packet (with inner and outer headers) should be created. Then it should be split into fragments below MAP size.

Initially, "tunnel MTU" and MAP could be manually synchronized by the administrator (with the difference in tunnel overhead). But any additional overhead on the tunnel path (nested tunnel, smallest Extension Header) would result in PMTUD that decreases MAP, but would not change "Tunnel MTU". It would turn on fragmentation for all bulk traffic. This situation is quite probable now (see [Huston-2020] on MTUs available on the Internet) and it would be even more probable in the future when many additional extension headers would be used. Hence, the requirement in section 5.3.1 "do NOT try to deprecate fragmentation" is indeed important.

Section 3.6 has the same approach as all other standards to the question of when fragmentation should happen: "this document assumes that only outer fragmentation is viable because it is the only approach that works for both IPv4 datagrams with DF=1 and IPv6". a considerable increase in fragmentation is proposed for the reasons

of academic purity: the router part of the router should behave as a router, the host part of the router should behave as a host without any deviations.

Additional fragmentations would create all of the problems discussed in [Fragile Fragmentation] and substantially increase the cost of tunnel endpoints. There is a high probability that draft [IP Tunnels] would be rejected by the market for cost reasons.

Additionally, we should point that statistics for fragmented packet drop on the Internet is still high enough (7%) - see [Huston-2021].

Fragmentation is the least probable solution for oversized packet drops.

3.4. PMTUD by original packet source

[PMTUD] is mandatory in IPv6 architecture, because IPv6 does not have fragmentation in transit. We could see recommendations in many RFCs not to block ICMPv6 PTB completely (it could be rate-limited - see [ICMPv6] section 2.4). [DPLPMTUD] section 1.1 has a very good collection of reasons why PTB message may not be delivered to the source - it is used as justification to augment PMTUD by [DPLPMTUD].

We should not see this problem for all non-tunneling protocols in the majority of environments. ICMPv6 PTB should be delivered to packet source, packet source would dynamically decrease PMTU to adapt to new realities. PMTU could change dynamically because some transit nodes could introduce additional extension header ad-hoc or ECMP could switch flow to a different path.

[IPv6 Tunneling] mandates to relay ICMPv6 PTB by tunnel ends for ICMPv6 messages received from the inside tunnel. [IPv6 Tunneling] does not use "relay" terminology, but section 8 explains in detail how to reconstruct and retransmit ICMP messages to the original packet source (delete all tunnel-related information). [MTU issues in Tunneling] section 3.2 discusses the same approach. [L2TPv3] section 4.1.4 refers to the [IPv6 Tunneling]. We could assume it as the request for PTB messages relay too. [SRH] section 5.4 confirms full adherence to ICMPv6 PTB relay approach: "For IP packets encapsulated in an outer IPv6 header, ICMP error handling is as defined in [IPv6 Tunneling]".

[VxLAN] section 4.3 proposes to use PMTUD: "Path MTU discovery MAY be used to address this requirement as well".

[NVO3] section 4.4.4 assumes PMTUD too: "It is strongly RECOMMENDED that Path MTU Discovery ([PMTUD]) be used to prevent or minimize

fragmentation".

[IPSec] section 8.2.1 requests that PMTU should be maintained for the tunnel and signaled to the original packet source as soon as any new packet would arrive.

[IPv6 GRE] section 3.3 clearly instructs developers to drop the oversized packets and send PTB for packets bigger than tunnel MTU. The method of PMTU detection is fully IPv6 compliant: "the GMTU is equal to the PMTU associated with the path between the GRE ingress and the GRE egress, minus the GRE overhead".

[MPLS Encapsulation] section 5.1 specifies the same approach: tunnel head-end should use [PMTUD] to understand tunnel MTU, then "the packet will have to be discarded, but the tunnel head should send the IP source of the discarded packet the proper ICMP error message".

[VxLAN], [NVO3], [IPSec], [IPv6 GRE], and [MPLS Encapsulation] do not request for tunnel endpoint to relay PTB messages. PMTUD should be used to set proper MTU for the tunnel, then subsequent packets could trigger PTB messages to the packet source. It would create an additional round trip delay compared to the original [IPv6 Tunneling] relay approach for the first PTB message. This small deficiency could be partially explained by the desire of many standards to be universal for IPv6 as well as IPv4. As a reminder, IPv4 may not have enough information in the ICMP message to properly reconstruct a relay message (64bits of source packet by RFC 792).

[IP Tunnels] is the only draft that contradicts [IPv6 Tunneling] (and every other protocol based on top) - it does clearly prohibit relay PTB messages. It states in section 3.3: "When such messages (PTB) arrive at the ingress interface ("ingress interface" is the tunnel interface in this draft), they may affect the properties of that interface (e.g., its MTU), but they should never directly cause new ICMPs in the outer network". This idea is generalized in section 5.1 as "ICMP messages MUST NOT be generated by the tunnel (as a link)". The motivation assumed in the draft is to fully mimic host behavior on the router virtual (tunnel) interface, because the host would not retranslate PTB messages.

We see that "Flow Label" is gaining popularity. [IPv6 Tunneling] and [ICMPv6] do not have strong recommendations for "Flow Label" - it was not an important topic at that time. The only small improvement that makes sense to do for [IPv6 Tunneling] is to recommend coping "Flow Label" from source packet to tunnel packet and from source packet to ICMPv6 PTB message. It would permit to properly load balance PTB messages to the same path as original traffic - see the problem [ICMPv6 PTB in ECMP] about hash-based load balancing between

many hosts. Copy "Flow Label" to PTB message would not contradict neither IPv6 architecture nor any RFC - it is not mandatory to develop a special standard update for it.

[MTU issues in Tunneling] section 3.2 has a concern that in the case of Lawful Intercept additional encapsulation could produce PTB messages that would show the fact to the monitored host. It is not a very realistic concern, because PMTU could change for many other reasons (especially with the proliferation of new protocols). If it is still a concern, then it makes sense to use another solution for this case: bigger MTU (better) or even fragmentation.

[MTU issues in Tunneling] section 3.2 raises the question about the applicability of "MSS Clamping". The transit node could snoop the transport layer and change MSS exchanged between nodes. This "hack" is not recommended because it breaks the layered model of IETF or OSI.

[PMTUD] is the only mechanism that is universal for all cases and fully compliant with IPv6 architecture. Vendors just need to use it, despite some challenges to relay PTB messages on tunnel ends. Moreover, it makes sense to standardize the relay of PTB messages on tunnel ends - it would improve PMTUD time on original traffic sources for round trip time.

[IPv6] RFC: "It is strongly recommended that IPv6 nodes implement Path MTU Discovery [PMTUD]".

3.5. Packetization Layer MTU Discovery

[PLPMTUD] and [DPLPMTUD] have been greatly developed in recent years. Packetization Layer (UDP/TCP) (1) has much more visibility (could see the size of transport layer buffers); (2) could operate under the absence of ICMPv6 PTB (too much filtering); (3) could be very granular (per-flow). It does have its use cases.

Albeit, PLPMTUD/DPLPMTUD have their restrictions as they: (1) are not universal for all transport protocols; (2) need more resources from the host; (3) are challenging to share PMTU information between applications; (4) need much more round trip times to find suitable PMTU; (5) do not work well on congested paths (difficult to understand the reason for packet loss).

Hence, PLPMTUD is not a replacement for PMTUD - both are needed. As a reminder from [PLPMTUD]: "Packetization Layer Path MTU Discovery (PLPMTUD) is most efficient when used in conjunction with the ICMP-based Path MTU Discovery".

PLPMTUD could play as a replacement for PMTUD in the worst-case scenario (ICMP is filtered). It would lead to the original host PMTU decrease too. PLPMTUD could be considered as a redundancy mechanism for PMTUD.

Strictly speaking, [PMTU by HbH] is a network layer mechanism, not a packetization layer. It is mentioned in this section because its usage is very similar to PLPMTUD, [PMTU by HbH] could be considered to some degree as the extension to PLPMTUD. It is not expected to principally change the conclusions of this document.

4. Conclusion

It is better not to have a problem with oversized packets in the first place. One should upgrade all links to a bigger MTU, if possible.

The host could have MTU as big as transit node. It would be never possible to deprecate PMTUD. It is important to follow the recommendations of [PMTUD] and [IPv6 Tunneling] for ICMPv6 PTB message delivery to the original traffic source. Tunnel sources should perform the relay function to make sure that the original traffic source would get the PTB message faster.

The temporary 220B limit for all headers pushes us to the frugal implementation of new extension headers. This limit would be alleviated after all backbone links would be upgraded to a much bigger MTU than 1500B. Additional protocols to collect MTU information could help in the transition period to attach additional headers frugally. It is true for all new protocols: SRv6, SFC, BIER, iOAM, APN6.

[PLPMTUD] and [DPLPMTUD] are not the replacement for [PMTUD], but could help in some scenarios.

Fragmentation is not at all a solution for oversized packet drops.

5. Security Considerations

[PMTUD], [PLPMTUD], [DPLPMTUD], and [Fragile Fragmentation] have some attack vectors discussed. This document does not introduce additional security vulnerabilities.

6. IANA Considerations

This document has no request to IANA.

7. References

7.1. Normative References

- [IPv6] S. Deering, R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 8200, DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.
- [ICMPv6] A. Conta, S. Deering, M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, DOI 10.17487/RFC4443, March 2006, <<https://www.rfc-editor.org/info/rfc4443>>.
- [PMTUD] J. McCann, S. Deering, J. Mogul, R. Hinden, "Path MTU Discovery for IP version 6", RFC 8201, DOI 10.17487/RFC8201, July 2017, <<https://www.rfc-editor.org/info/rfc8201>>.
- [IPv6 Tunneling] A. Conta, S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, DOI 10.17487/RFC2473, December 1998, <<https://www.rfc-editor.org/info/rfc2473>>.
- [ICMPv6 PTB in ECMP] M. Byerly, M. Hite, J. Jaeggli, "Close Encounters of the ICMP Type 2 Kind", RFC 7690, DOI 10.17487/RFC7690, January 2016, <<https://www.rfc-editor.org/info/rfc7690>>.
- [MTU issues in Tunneling] P. Savola, "MTU and Fragmentation Issues with In-the-Network Tunneling", RFC 4459, DOI 10.17487/RFC4459, April 2006, <<https://www.rfc-editor.org/info/rfc4459>>.
- [IP Tunnels] J. Touch, M. Townsley, "IP Tunnels in the Internet Architecture", draft-ietf-intarea-tunnels-10 (work in progress), September 2019.
- [IP Encapsulation] C. Perkins, "IP Encapsulation within IP", RFC 2003, DOI 10.17487/RFC2003, October 1996, <<https://www.rfc-editor.org/info/rfc2003>>.
- [IPSec] S. Kent, K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <<https://www.rfc-editor.org/info/rfc4301>>.

- [IPv6 GRE] C. Pignataro, R. Bonica, S. Krishnan, "IPv6 Support for Generic Routing Encapsulation (GRE)", RFC 7676, DOI 10.17487/RFC7676, October 2015, <<https://www.rfc-editor.org/info/rfc7676>>.
- [MPLS Encapsulation] T. Worster, Y. Rekhter, E. Rosen, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", RFC 4023, DOI 10.17487/RFC4023, March 2005, <<https://www.rfc-editor.org/info/rfc4023>>.
- [L2TPv3] J. Lau, M. Townsley, I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, DOI 10.17487/RFC3931, March 2005, <<https://www.rfc-editor.org/info/rfc3931>>.
- [VxLAN] M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <<https://www.rfc-editor.org/info/rfc7348>>.
- [NVO3] J. Gross, I. Ganga, T. Sridhar, "Geneve: Generic Network Virtualization Encapsulation", RFC 8926, DOI 10.17487/RFC8926, November 2020, <<https://www.rfc-editor.org/info/rfc8926>>.
- [L3VPN] E. Rosen, Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <<https://www.rfc-editor.org/info/rfc4364>>.
- [EVPN] A. Sajassi, R. Aggarwal, N. Bitar, A. Isaac, J. Uttaro, J. Drake, W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <<https://www.rfc-editor.org/info/rfc7432>>.
- [Huston-2016] Huston, G., "Fragmenting IPv6", Blog Post, 2016, <<https://blog.apnic.net/2016/05/19/fragmenting-ipv6/>>.
- [Huston-2021] Huston, G., "IPv6 Fragmentation Loss", Article, 2021, <<https://www.potaroo.net/ispcol/2021-04/v6frag.html>>.
- [Fragile Fragmentation] R. Bonica, F. Baker, G. Huston, R. Hinden, O. Troan, F. Gont, "IP Fragmentation Considered Fragile", RFC 8900, DOI 10.17487/RFC8900, September 2020, <<https://www.rfc-editor.org/info/rfc8900>>.

7.2. Informative References

- [PLPMTUD] M. Mathis, J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, DOI 10.17487/RFC4821, March 2007, <<https://www.rfc-editor.org/info/rfc4821>>.
- [DPLPMTUD] G. Fairhurst, T. Jones, M. Tuexen, I. Ruengeler, T. Voelker, "Packetization Layer Path MTU Discovery for Datagram Transports", RFC 8899, DOI 10.17487/RFC8899, March 2020, <<https://www.rfc-editor.org/info/rfc8899>>.
- [SRH] C. Filsfils, D. Dukes, S. Previdi, J. Leddy, S. Matsushima, D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.
- [PMTU by HbH] R. Hinden, G. Fairhurst, "IPv6 Minimum Path MTU Hop-by-Hop Option", draft-hinden-6man-mtu-option-02 (work in progress), July 2019.
- [PMTU by ISIS] Z. Hu, Y. Zhu, Z. Li, L. Dai, "IS-IS Extensions for Path MTU", draft-hu-lsr-isis-path-mtu-00 (work in progress), June 2018.
- [PMTU by PCEP] S. Peng, C. Li, L. Han, "Support for Path MTU (PMTU) in the Path Computation Element (PCE) communication Protocol (PCEP)", draft-li-pce-pcep-pmtu-03 (work in progress), October 2020.
- [PMTU by BGP-LS] Y. Zhu, Z. Hu, G. Yan, J. Yao, "BGP-LS Extensions for Advertising Path MTU", draft-zhu-idr-bgp-ls-path-mtu-05 (work in progress), November 2020.
- [PMTU by SR-Policy] C. Li, Y. Zhu, A. Sawaf, Z. Li, "Segment Routing Path MTU in BGP", draft-li-idr-sr-policy-path-mtu-03 (work in progress), November 2019.
- [Generic Delivery Functions] Z. Zhang, R. Bonica, K. Kompella, "Generic Delivery Functions", draft-zzhang-intarea-generic-delivery-functions-01 (work in progress), April 2021.
- [Pseudowire Fragmentation] A. Malis, M. Townsley, "Pseudowire Emulation Edge-to-Edge (PWE3) Fragmentation and Reassembly", RFC 4623, DOI 10.17487/RFC4623, August 2006, <<https://www.rfc-editor.org/info/rfc4623>>.

8. Acknowledgments

Thanks to v6ops working group for problem discussion

Authors' Addresses

Eduard Vasilenko
Huawei Technologies
17/4 Krylatskaya st, Moscow, Russia 121614

Email: Vasilenko.Eduard@huawei.com

Xiao Xipeng
Huawei Technologies
205 Hansaallee, 40549 Dusseldorf, Germany

Email: Xipengxiao@huawei.com

Dmitriy Khaustov
Rostelecom
13/2 Nikoloyamskaya st, Moscow, Russia 109240

Email: Dmitriy.Khaustov@rt.ru

