Historic Developments in IPv6 Measurements

Florian Briksa, Lion Steger*
*Chair of Network Architectures and Services
School of Computation, Information and Technology, Technical University of Munich, Germany
Email: florian.briksa@tum.de, stegerl@net.in.tum.de

***Abstract*—Over the last decades, the Internet has become an important and integral part of our daily lives. Handling increasingly large amounts of devices interacting over the Internet, the old address space of Internet Protocol Version 4 (IPv4) is becoming too small. Therefore, in 1998 Internet Protocol Version 6 (IPv6) was developed, which has a significantly larger range of addresses. But IPv6 has also some disadvantages such as readability and, especially important for Internet Providers, it is almost impossible to keep track of all addresses being used on the Internet. This makes lists of currently active addresses, called hitlists, necessary. This paper analyses the growth process of IPv6 along with outliers in the data provided by a selected hitlist and offers a detailed view into composition of aliased addresses and usages of IPv6 around the world.**

***Index Terms*—IPv6, measurement, historical analysis, aliasing**

1. Motivation

As IPv6 prevalence continues to grow, analyzing its usage is becoming increasingly important. Network administrators want to know traffic origins, while Internet Service Providers want to reliably allocate addresses and deliver information to their customers. Moreover, observant analysts can detect shifts in IPv6 usage during significant events like wars or disasters. These scenarios underscore the necessity for robust tools to analyze trends and detect anomalies in the IPv6 address space.

This paper focuses on fundamental analyses of IPv6 address space development. Section 5 delves into the composition of prefixes used for addresses and the countries utilizing IPv6 from 2018 to 2024, using a hitlist maintained by the Chair of Network Architectures and Services since 2018 [1] and geolocation tools. It presents a comprehensive view of the address space growth and identifies countries which have the biggest impact on communication over IPv6 according to the referred data.

In Section 6, we further explore outliers in the data, providing a concise before-and-after summary of address space changes and discussing potential origins.

2. Related Work

This paper is based on data obtained from the IPv6 hitlist maintained by the Chair of Network Architectures and Services since 2018 [1]. This hitlist is in the following just referred to as "hitlist". It also utilizes geolocation and technical information from the International Assigned Numbers Authority (IANA) [2], which oversees the assignment and usage of all IPv6 addresses assigned to customers worldwide. The works by Gasser et al. [3], Zirngibl et al. [4], and Steger et al. [5] were particularly helpful in identifying outliers resulting from internal changes in scan execution methodologies.

For comparison between geolocation tools and overall IPv6 usage in different countries, the insights provided by APNIC Labs [6] provided suitable information, particularly in terms of IPv6-capable and IPv6-preferring devices.

3. Methodology

During our research, we developed a tool to process hitlist information in multiple aspects. It was used to filter and create diagrams used in the following, along with a database interface for more efficient processing. To locate the country of IP-addresses, we used the WHOIS [7] database by the Internet Assigned Numbers Authority (IANA) [2].

As in this paper we focus more on countries than on exact addresses, we can reduce the lookups to WHOIS by storing the first 32 bits of each address in a database. This is possible because an ISP typically gets assigned the first 32 or fewer bits of an IPv6 address space for its customers from a Regional Internet Registry (RIR). It can now be assumed that most of the companies or institutions using those addresses operate in their home country, which makes the country identification up to 99% accurate [8].

It is important to mention that the WHOIS database only provides information about the country an AS is assigned to, not the servers on which the AS is running. Therefore, the precision of assigned and operating country may vary.
It also has to be noted that, as an exhaustive scan over all IPv6 addresses is not possible, the results presented in this paper may vary across different hitlist generators, as they possibly have completely different or varying generation methods of finding addresses [5].

4. Background

At first, we provide some background information to offer a clearer view of the research in this paper.

4.1. IPv6 Hitlist

The hitlist used in this paper is maintained by the Chair of Network Architectures and Services and includes

 doi: 10.2313/NET-2024-09-1_08

lists categorized by aliased and non-aliased addresses, along with lists categorized by used protocols. The entries of the hitlist contain IPv6 addresses of responding servers during a regular scan of the address space. However, this paper primarily focuses on aliased prefixes because they illustrate the structure of IPv6 [9] addresses used in networks and allow a more efficient analysis, although they do not hold as much information as non-aliased addresses, as described in Chapter 4.4.

4.2. IPv6 Notation and Prefix

Each IPv6 address consists of 128 bits available for address location, providing a larger address space than IPv4 which has 32 bits available. It is followed by a number representing the prefix length of this address in bits. For example, an entry could look like this:

```
2401:4900:22dc:fab9::/64          (1)
```

Here we see a common IPv6 address in the usual CIDR (Classless Inter Domain Routing [10]) notation. The number after the slash describes the length of the prefix, in this case, 64 bits. The prefix is used for dividing the address space into sub address spaces of variable size. The longer the prefix, the smaller the resulting sub-address space. In this 64-bit prefix example, we would have 64 bits left to choose addresses for our devices in our network.

4.3. Network Categories and Protocols

The data used contains addresses from diverse network categories, including ISP (Internet Service Provider), NSP (Network Service Provider) and CDN (Content Delivery Network). These categories can be assigned by network operators to their Autonomous System (AS). As described in the paper by Lion Steger et al. [5], more than 42% of hitlist addresses are allocated to ISP networks. Furthermore, this paper considers the distinct behavior of addresses associated with their respective network categories and analyzes possible correlations between protocol and AS/Prefix composition.

Devices communicating over IPv6 use multiple protocols for message transmission. These protocols are the key to measuring the responsiveness of addresses. The hitlist used in our research conducts scans for TCP/80 (HTTP) and TCP/443 (HTTPS), ICMP, UDP/53 (DNS) and UDP/443 (QUIC) on a regular basis [5].

The last protocol to mention here is the ”Internet Control Message Protocol for the Internet Protocol Version 6” (ICMPv6). ICMPv6 is an important part of communicating with IPv6, as it reports errors and provides diagnostics [11], and all parts of this base protocol have to be implemented in all nodes communicating over IPv6.

4.4. Aliasing

IPv6 addresses can be further divided into aliased and non-aliased addresses. Gasser et al. [3] described aliased prefixes as subnets where every address in this subnet is mapped to and responded to by one single host, identified by this aliased prefix. Therefore, the number of aliased prefixes is usually much smaller than non-aliased ones.

However, aliased addresses usually do not hold as much information as non-aliased addresses, as they are used by ASes and not by single devices [3].

5. Basic Analysis

In this chapter we focus on long term trends visible in our processed data. For now, we ignore bigger outliers as much as possible to obtain a better view of the overall development of IPv6 usage in recent years.

5.1. Analysis of AS/Prefix Composition

In this first subchapter we start analysing the prefix composition of responsive addresses from July 2018 to april 2024. In Figure 1 we can observe the development of AS/Prefix composition.

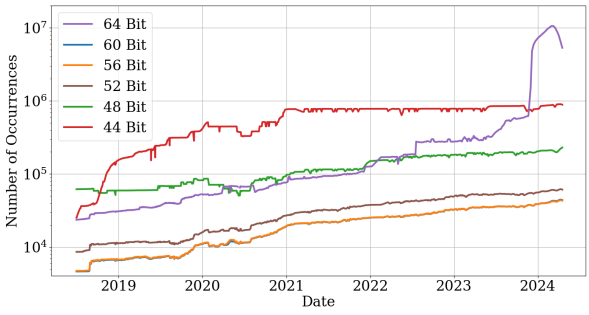

Figure 1: Composition of the six most used prefix lengths from year 2018 to 2024

At first, we can clearly outline that from 35 different prefix lengths from 29 to 120 bits, only three are extensively found: 44 bits, 48 bits and 64 bits. For end users, the use of 64-bit for addresses is recommended, as it is required for Stateless Address Autoconfiguration (SLAAC) to work [12]. SLAAC is used to generate IPv6 addresses for devices in a network without any further control from outside. The usage for end users is evident, as all of our most evaluated prefix lengths are at least 64 bits long. There may be several reasons to further divide the given address space into smaller subnets. One reason could be that an end user wants to connect multiple devices under the same prefix or uses several virtual machines in their network, having an individual internal routing topology. Another reason might be a network plan that is easier to remember. The host gets the original 64-bit prefix address, and then hierarchically structured sub-devices get the next 8 or 16 bits of their corresponding subnet etc [12].

Figure 1 also shows that from the beginning of the measurements in 2018 until July 2022 the 44 bit prefix clearly dominated and rose, while all other prefix lengths remained mostly stable. After January 2021, the number 44 bit prefixes found remained stable. This may be the result of new addresses being added under the already existing prefixes, and therefore not being added to the hitlist. In July 2022, the hitlist added new address candidates from new passive sources and target-generation methods [4]. This led to a general rise in the number

of responsive addresses logged as well as in different categories of IPv6 addresses, especially in ISP (Internet Service Provider), NSP (Network Service Provider) and CDN (Content Delivery Network). It follows that, as we only have a rise in 64-bit prefixes at the same time, those categories directly correlate with our change in prefix length composition.

To conclude, it can be noted that not surprisingly most of the devices in the IPv6 address space are likely end users. Furthermore, it is possible that end users divide their address space into smaller subnets to better organize themselves.

5.2. Analysis of Geolocation Data

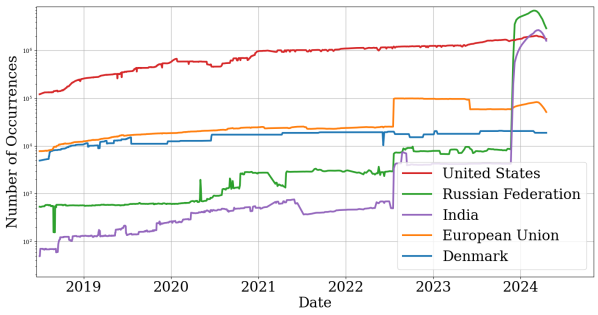

Figure 2: Composition of the six most active Countries from 2018 to 2024

Upon initial examination of Figure 2, it is clear that the United States has the most active and responsive IPv6 addresses, while addresses from outside the US did not play a significant role until July 2022. As mentioned earlier, new methods for finding IP addresses were applied to the hitlist at this date. Following July 2022, we can observe a visible but still relatively small increase in addresses originating from within the European Union.

Up to this point, the found addresses might indicate a lower priority of IPv6 in most countries except for the US. Only in recent months we have noted a slight increase in addresses originating from Vietnam. Upon analyzing the addresses from the EU, Denmark has, according to our data, had the most active addresses found in recent years.

It is worth mentioning that the most active addresses come from the US, India, the EU, and Russia. The only country not in the top five most active countries is China, which is not even noticeable among the other smaller countries in the diagram. This may be a result of the organized censorship of foreign servers under the Great Firewall of China, leading to only a few servers being connected to the rest of the world [5]. As Zirngibl et al. describes, those addresses lead to peaks and inaccuracies in the data. Therefore, most of the Chinese addresses are filtered [4].

We can observe a correlation between AS/Prefix Composition and geolocation data. The increase in 44-bit prefixes, followed by 48- and 64-bit prefixes, corresponds with the growing number of addresses originating from the US.

6. Analysis of Outliers

In this chapter, we deal with the identification of outliers in the examined data and further try to analyze their origins.

6.1. Jumps in Hitlist Data

The way addresses are scanned has a great impact on number and composition. For example, in the following section we describe a jump that occurred in July 2022. During this period of time, the found addresses with 64-bit prefixes increased from 186,000 to 277,000. The possible reason for this may be the paper published by Zirngibl et al. [4] in 2022, which presented new address candidate sources along with target-generation algorithms, the scans of the address space found more aliased addresses especially with 64 bit prefixes. This correlates with a jump in the overall number of responsive addresses found over ICMPv6.

Such jumps are not uncommon, as changes in algorithms are continuously applied and offer a wide research area. On the other hand, sudden breakouts may also happen when networks with greater numbers of addresses block parts of their address spaces from access from outside, making addresses unresponsive and therefore not listed in the data [1].

6.2. Plunge in Responsive Addresses in the US

The first outlier in our data happened in 2020 and was already visible in the previous figures:

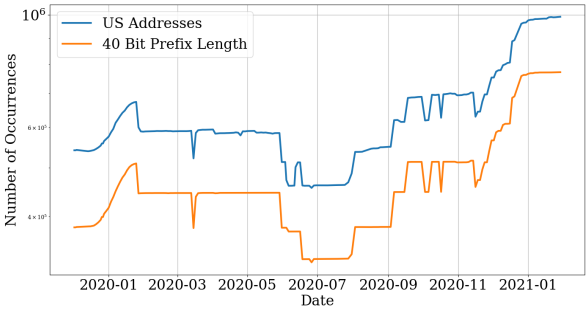

Figure 3: Outlier in Prefix Composition and US Addresses 2020

As prefix composition and number of responsive addresses from the US are clearly connected, we can assume that events inside the US were responsible for that outlier.

Before 2020, we found increasing numbers of responsive addresses originating inside the US. When we have a look at the data, the number of addresses had increased to 670,000 active addresses on January 24th, 2020, and sunk to 460,000 on June 24th.

After November 2020, the number of aliased addresses rapidly increased beyond the number measured before January, reaching a new maximum of one million aliased addresses.

The evaluation of more than 90% of US addresses leads to AWS Cloud Services in Seattle. As [13] states,

more than 40 states used at least one of Amazon's election services in the 2020 presidential election campaigns.

As it is quite common for admins to block frequent address scans for hitlists in their firewalls, it may be possible that AWS has blocked addresses to critical infrastructure during the election period, causing this outlier in responding addresses from the US [3]. As AWS uses over 97% of the entire addresses located inside the US, we get a plunge during that election period in 2020.

6.3. Recent Peak in Responsive Addresses

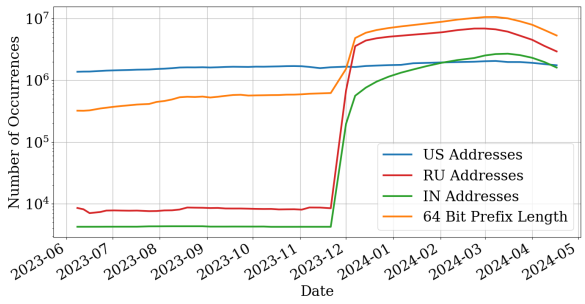

Figure 4: Outlier in Prefix Composition and Russian/Indian Addresses 2024

In the second major outlier to be discussed in this paper we can observe an immensely increasing number of 64-bit addresses especially originating from the Russian Federation and India, surpassing the number of addresses found in the US.

Before of this peak, the scans found fewer than 10,000 aliased addresses originated in Russia and India. During the following three months, responsive addresses from Russia increased to over 6.8 million and India to 2.7 million, while the dominating country in our scans, the United States, continuously increased to 2.1 million.

During our research, this number decreased as abruptly as it increased three months before. As visible in Figure 4, Russia and India still had the majority of scanned addresses, but a much lower level than at its maximum with 2.9 million and 1.6 million responsive addresses, respectively.

This peak may be the result of applying new filters and target-generation algorithms used by scanners to find active IPv6 addresses during this period. As the numbers also decreased at the same rate, we can assume that network administrators have blocked more addresses from being pinged by scans. As multiple Russian servers decreased their responsive addresses at a similar rate, it is possible that those servers have the same administrators, applying filters for their firewalls at the same time [3] [5].

Another reason may be the re-evaluation of addresses after being unresponsive for 30 days, causing a significant increase in protocol responsiveness over ICMPv6 as displayed as event "I" in Figure 5 [1].

7. Conclusion and Future Work

In this paper, we discussed and analyzed various outliers in IPv6 hitlist data. We provided an overview of

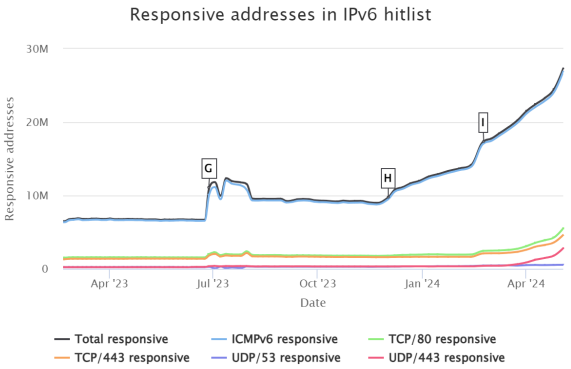

Figure 5: Protocol Responsiveness from January 2023 until now [1]

the overall development of IPv6 usage in recent years and explained the methodology behind collecting and structuring hitlist data.

We can clearly observe the general trends in IPv6 development, although the aliased addresses used for this paper represent only a small portion of globally active and responsive addresses. However, this subset of responsive addresses already provides insights that can help draw conclusions related to specific events in the countries where they occurred. We discovered that Chinese addresses have mostly been blocked, resulting in their underrepresentation in our dataset. On the other hand, India has significantly expanded and modernized its communication infrastructure, leading to a notable increase in responsive addresses.

In future research, it would be beneficial to extend this analysis to non-aliased addresses. Additionally, further analysis of the geographical locations of address origins using the implemented analysis tool could yield valuable insights. By comparing addresses located in different countries with global IPv6 usage statistics, we can draw conclusions about the composition and development of internet service infrastructure in those countries.

References

[1] O. Gasser, J. Zirngibl, and L. Steger:, "IPv6 Hitlist," https://ipv6hitlist.github.io, 2024, [Online; accessed 5-May-2024].

[2] I. A. N. Authority, "IANA," https://www.iana.org/whois, 2024, [Online; accessed 12-May-2024].

[3] O. Gasser, Q. Scheitle, P. Foremski, Q. Lone, M. Korczynski, S. D. Strowes, L. Hendriks, and G. Carle, "Clusters in the expanse: Understanding and unbiasing ipv6 hitlists," 2018.

[4] J. Zirngibl, L. Steger, P. Sattler, O. Gasser, and G. Carle, "Rusty clusters? dusting an ipv6 research foundation," 2022.

[5] L. Steger, L. Kuang, J. Zirngibl, G. Carle, and O. Gasser, "Target acquired? evaluating target generation algorithms for ipv6," Jun. 2023.

[6] A. Labs, "IPv6 Preferred Rate by country," https://stats.labs.apnic.net/ipv6/, 2024, [Online; accessed 14-May-2024].

[7] IANA, "IANA WHOIS," https://www.iana.org/whois, 2022, [Online; accessed 14-May-2024].

[8] iplocation.net, "How accurate is IP-based Geolocation Lookup?" https://www.iplocation.net/geolocation-accuracy, 2016, [Online; accessed 12-May-2024].

[9] IANA, "IPv6 Specification," https://www.rfc-editor.org/rfc/rfc8200.html, 2017, [Online; accessed 14-May-2024].

[10] AWS, “Was ist CIDR?” https://aws.amazon.com/de/what-is/cidr/, [Online; accessed 28-May-2024].

[11] T. I. Society, “Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification,” https://datatracker.ietf.org/doc/html/rfc4443, 2006, [Online; accessed 14-May-2024].

[12] serverfault.com, “How does IPv6 Subnetting Work and How does it differ from IPv4 Subnetting?” https://serverfault.com/questions/426183/how-does-ipv6-subnetting-work-and-how-does-it-differ-from-ipv4-subnetting, 2022, [Online; accessed 14-May-2024].

[13] Reuters:, “How Amazon moved into the business of U.S. elections,” https://www.nbcnews.com/tech/tech-news/how-amazon-moved-business-u-s-elections-n1066286, 2020, [Online; accessed 15-May-2024].