Evaluation of sources for IPv6 Hitlists

Finn Johannes Hartmann, Lion Steger*
*Chair of Network Architectures and Services
School of Computation, Information and Technology, Technical University of Munich, Germany
Email: finn.hartmann@tum.de, stegerl@net.in.tum.de

***Abstract*—IPv6 hitlists are essential tools for conducting large-scale internet measurements. This paper evaluates various data sources used to construct IPv6 hitlists, focusing on IPv6 address stability and Border Gateway Protocol (BGP) distribution. Our analysis reveals significant differences in the quality and characteristics of each source's data and highlights the importance of source selection in measurement studies.**

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

1. Introduction

The ongoing adoption of IPv6 has introduced many challenges due to the large scale of the address space (2^{128} addresses). Due to this size, it is infeasible to scan the whole space. To address this need, the 'IPv6 Hitlist Service' maintained by the Chair of Network Architectures and Services at TUM [1] provides a continuously updated list of responsive IPv6 addresses. In the following, we will refer to it as 'hitlist'. The hitlist contains IPv6 addresses from different public sources and verifies their responsiveness regularly. Hence, it has become an essential tool for internet scans.

Previous work by Steger et al. [2] established this hitlist methodology, while Gudehege [3] analyzed the source quality deeply.

In this paper, we will analyze such an IPv6 hitlist, and focus on two other aspects:

- The stability of Autonomous System Numbers (ASNs) [4] for IPv6 addresses over time (Section 4.1)
- The source distribution across ASNs (Section 4.2)

With these analyses, we try to find a choice of sources that will improve future measurement efforts in their overall representativeness. We aim to inform future measurement efforts by better understanding the relationship between sources and ASNs and address stability.

2. Data Overview

This section provides foundational information on the hitlist dataset and the methods used to collect and process the data. The subsequent analysis presented in Section 4 is based on this background.

We are using a dataset collected by the IPv6 Hitlist Service [1]. The hitlist contains a filtered list of reachable IPv6 addresses gathered from their last scan [5]. To perform these scans, they use various sources, e.g., the hitlist, Bitnodes, and IPInfo (Section 2.2). Hence, the hitlist shows IPv6 addresses from the active Internet address space and is updated regularly, nowadays on a monthly basis, but in the earlier days weakly or daily, during each scan.

These addresses are subsequently categorized based on their responsiveness to different network protocols, enabling a deeper understanding of IPv6 address space utilization [2].

For our evaluation, we worked on an SQL import of the hitlist by Gudehege [3], which is structured into two main components:

- **Inputdata**: All collected IPv6 addresses, irrespective of their responsiveness, are active unless our scan reverts this assumption.
- **Outputdata**: A filtered subset of the inputdata that includes only addresses that responded during active scans.

2.1. Data Structure

The distinction between the inputdata and outputdata tables is essential for our analysis. Not all relevant metadata is presented in both tables, necessitating the use of both to perform a comprehensive evaluation.

- **Inputdata Table Fields**: `ip`, `source`, `date`, `bgpIp`, `bgpMask`, `asn`
- **Outputdata Table Fields**: `ip`, `protocol`, `date`, `bgpIp`, `bgpMask`, `asn`

The key difference lies in the `protocol` field in the outputdata and the fact that only responsive IPv6 addresses are retained. This makes the outputdata particularly useful for filtering reachable IPs.

2.2. Data Collection

The IPv6 addresses were aggregated from a diverse range of publicly available sources, including e.g.:

- Bitnodes — provides active Bitcoin nodes, which operate over IPv6 [6]
- IPInfo — offers a dataset which contains enriched metadata for their observed IPs, including e.g. geolocations and ASN [7]
- Rapid7DNS — contains IPv6 responses from large-scale zone file scans [8]
- RIPE — collects data by using globally distributed probes, which are based on a community supported measurement [9]

 doi: 10.2313/NET-2025-11-3_15

- Hitlist — is maintained by the Chair of Network Architectures and Services which combine all these different sources and perform their scans on this data to update their list of responsive IPv6 addresses and performance DNS resolution [1], [5]

This comprehensive aggregation strategy ensures broad coverage of the active IPv6 address space. The database on which our analyses are based contains approximately 11B entries. Therefore, we limited the data volume for our evaluation by using a sampling condition in which we only included every 100th entry.

3. Related work

This paper is based on data collection from the IPv6 hitlist maintained by the Chair of Network Architectures and Services since 2018 by Gasser et al. [1]. They have introduced a methodology for generating IPv6 hitlists by aggregating data from various sources. The hitlist structure on which we are working is based on the SQL structure from Gudehege [3]. This structure is explained in the background (Section 2.1).

Our work extends these studies by providing a comparative analysis of data sources concerning IPv6 address stability and BGP [10] distribution, offering insights into the suitability of each source for different measurement objectives.

4. Analysis

During the analysis in this chapter, we focus on two main aspects: IPv6 address stability (Section 4.1) and source ratio per ASN (Section 4.2). The analysis is based on measurements from the `hitlistdb.inputdata` dataset. Each source is analyzed over time to detect volatility patterns and systematic biases in network distribution.

4.1. Analysis of ASN stability

This subsection examines, we examine the ASN changes from November 2023 to October 2024, focusing on the temporal patterns and source-specific characteristics of IPv6 address stability.

4.1.1. Data preparation. In the following analysis, we work with a subset of our inputdata, specifically querying the dataset for all IPv6 addresses that have appeared with more than one Autonomous System Number (ASN). From this query, we gather a list of 'unstable' IPs exhibiting ASN changes, which were used for a detailed timeline visualizations of ASN stability patterns.

4.1.2. Observations. Figure 1 shows the monthly count of IPv6 addresses that changed their ASN, broken down by data source. Additionally, Figure 2 presents the relative fraction of IPs with ASN changes per source, providing normalized insights into source-specific volatility patterns.

Monthly ASN Change Patterns: The monthly ASN change analysis reveals significant temporal and source-specific variations in IPv6 address stability. From November 2023 to February 2024, we observe a noticeable peak in the total number of ASN changes in IPs, reaching approximately 2,500–3,000 per month. Afterward, the total number of IPs with ASN changes declines and stabilizes around 1,000–1,500 changes per month. Such a noticeable difference could be caused by increased network infrastructure changes or BGP hijacking [11]. The latter refers to a falsely announced BGP prefix by an Autonomous System (AS).

Source-Specific Volatility Patterns: The clustering of sources shows a clear behaviour pattern across different sources. High-volatility sources (including *ipinfo*, *yarrp*, and *rapid7dns*) contribute disproportionately to ASN changes throughout the observation period. However, such a high proportion could also be based on the size of these sources (TABLE 1), which is, in our case, very likely. In peak months, *ipinfo* and *rapid7dns* show a similar pattern to the total cumulation. Such a pattern indicates that these sources capture IP addresses that are associated with dynamic or ephemeral infrastructure. This leads to the effect that they are assigned only temporarily and change more frequently. Each newly assigned IPv6 address may also map to a different Autonomous System due to load balancing or routing optimization. Against that, there are stable sources such as *bitnodes*, *openipmap*, and *ripe*, where assigned IP addresses infrequently change over time. These sources rarely exceed 200–300 changed ASNs monthly, making them more suitable for long-term network topology studies.

Source	Count
ipinfo	8.168.647.739
hitlist	2.349.036.709
yarrp	541.657.853
rapid7dns	228.492.805
ripe	202.662.048
bitnodes	50.812.377
openipmap	45.478.545

TABLE 1: IPv6 address counts per data source

However, unlike the total number of IP changes, the relative numbers show that *ipinfo* and *yarrp* have consistent IP changes. The relative numbers are based on the total number of IP changes for the corresponding source. More than 50% of *ipinfo* and more than 40% of *yarrp* IPs change their ASNs monthly. Generally, there is nearly the same pattern as shown in Figure 1. Initially, all sources have a high change rate, spreading from 10% to 100%. From March to October, we can observe clusters. One has a high change rate of roughly 50%–90%, and the other from around 0%–45%.

Temporal Stability Trends: The data shows an overall decline in ASN changes with a constant decrease from the peak in early 2024 to approximately 1,000–1,500 per month.

Figures 3 and 4 illustrate the ASN change patterns per IP over time by plotting temporal ASN assignment intervals for individual IP addresses.

The frequency of ASN changes observed for specific IP addresses, particularly those shown in Figure 3, represent an uncommon behavior in typical internet infrastructure. The analysis reveals that most IPv6 addresse with frequent ASN share a common BPG prefix. Two

 doi: 10.2313/NET-2025-11-3_15

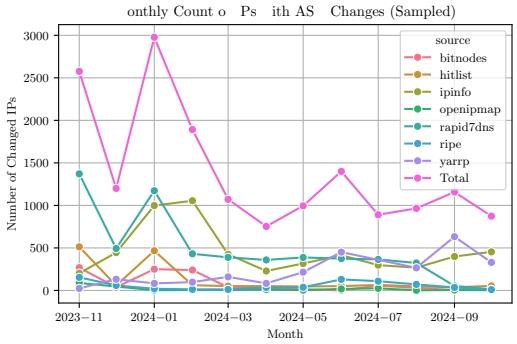

Figure 1: Monthly ASN changes per source

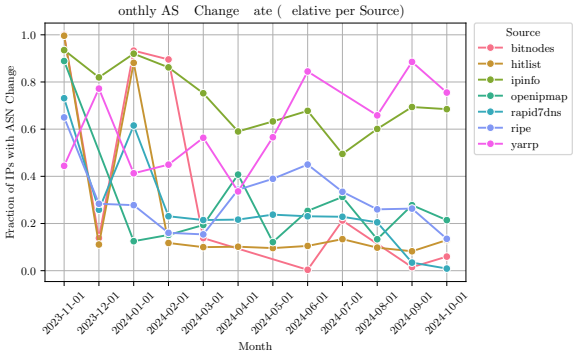

Figure 2: Monthly ASN change rate (relative per source)

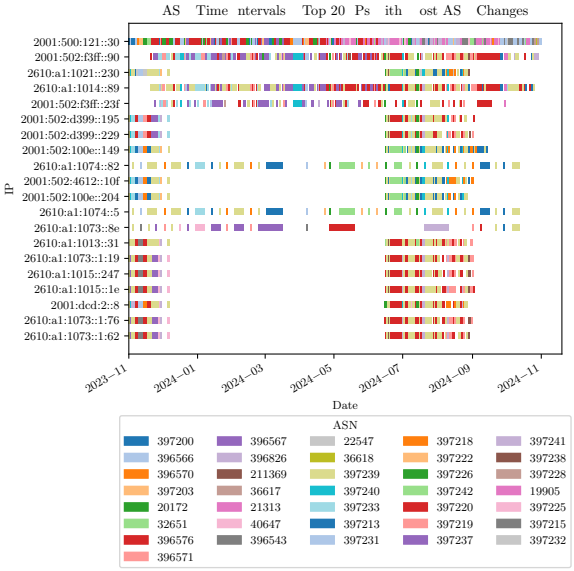

Figure 3: Top 20 IPv6 addresses with most ASN changes

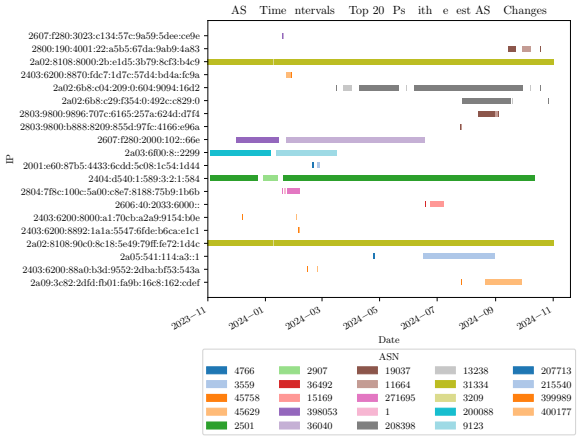

Figure 4: Top 20 IPv6 addresses with the fewest non-zero ASN changes

representative examples include:

`2001:500:121::30` (1)

`2001:502:f3ff::90` (2)

This behavior is most likely attributable to Multi-Origin AS (MOAS) prefixes, where the same IPv6 BGP prefix is announced by multiple ASNs simultaneously or over time [12]. This phenomenon serves as a strong indicator for dynamic, ephemeral infrastructure deployment patterns or as a way of load balancing on the AS-level (`2001:500:121::/48` or `2001:502:f3ff::/48` [13], [14]).

These addresses switch ASNs multiple times between 2023 and 2024, including changes to AS397239 (Vercara, LLC), AS397220 (Vercara, LLC), and AS396566 (VeriSign Global Registry Services) [15]. By often using the same ASNs, we can strongly assume that these addresses are used temporarily.

Conversely, Figure 4 presents IPv6 addresses with stable ASN assignments throughout the observation window. Examples include:

`2a02:8108:8000:2b:e1d5:3b79:8cf3:b4c9` (3)

`2404:d540:1:589:3:2:1:584` (4)

These findings emphasize the need for stability-aware selection when analyzing IPv6 topologies. Ignoring ASN volatility may lead to biased assumptions about the persistence or reachability of observed prefixes. Using diversified and stable sources, as recommended in [2], [3], [12], strengthen measurement accuracy.

4.2. Source distribution per ASN

During this chapter, we analyze the source distribution per ASN for the 10 ASNs with the most entries in the hitlist.

 doi: 10.2313/NET-2025-11-3_15

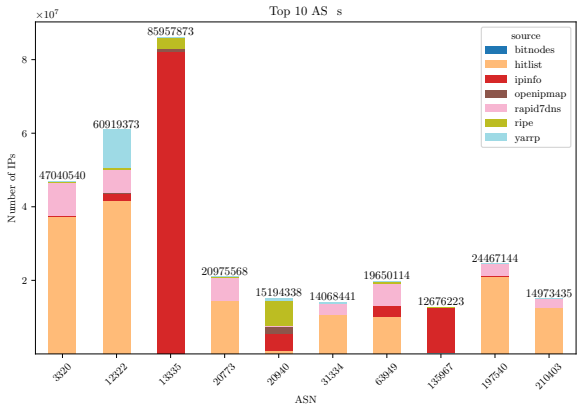

Figure 5: Top 10 ASNs with corresponding source contributions

4.2.1. Data preparation. To analyze the distribution of IPv6 addresses across ASNs by source, we extracted all records containing valid BGP prefix information and ASN. We grouped entries by BGP prefix, ASN, and source. Subsequently, we counted the number of IPs associated with each combination. If one IPv6 address had more than one source, it was counted for every source individually. Hence, we explicitly considered from which source an IPv6 address was imported. Finally, we identified the top 10 ASNs with the highest IP counts and visualized the source contributions, which are shown in a stacked bar plot.

4.2.2. Observations. Figure 5 reveals significant imbalances in source contributions across the top 10 ASNs by IP count, with entry counts ranging from approximately $1.2{\times}10^7$ to $8.6{\times}10^7$ across different ASNs.

ASN Distribution Imbalances: An essential observation from the distribution is the pronounced source dominance pattern. AS13335 (Cloudflare) shows the highest total count (approximately $8.6{\times}10^7$ entries) with dominant contributions from *ipinfo* and *hitlist* sources.

Representativeness Implications: The significant imbalances cause several concerns for IPv6 measurement studies — Geographic bias from single-source dominance may skew results toward specific regions, source-specific collection methodologies may introduce systematic temporal biases, and certain sources may preferentially capture specific infrastructure types (e.g., CDN nodes, residential addresses, or enterprise networks) [3], [16].

Certain ASNs, such as AS13335 (Cloudflare Inc.), AS3320 (Deutsche Telekom AG), and AS12322 (Free SAS) [15], are heavily dominated by data from one or two sources, e.g., *ipinfo* or *hitlist*. This skew can result in an overrepresentation of specific network regions, which may bias scanning results if only single-source hitlists are used.

Hence, incorporating multiple diverse sources is crucial to improve ASN diversity and global representativeness in IPv6 measurements [3].

5. Conclusion and Future Work

After analyzing ASN stability and source distribution per ASN, the evaluation of IPv6 hitlist entries indicates significant stability of IPv6 addresses and variability of BGP distribution. Sources like Bitnodes and Yarrp provide more stability, making them preferable for longitudinal studies. In contrast, while offering broader coverage, sources such as IPInfo and Rapid7DNS may introduce volatility and skewed ASN representation.

Recommendations for Future Measurements: Based on our analyses and the distribution patterns observed, we recommend multi-source integration to mitigate single-source biases, source-aware sampling with weighted contributions to achieve balanced ASN representation, stability-informed selection, and prioritizing stable sources for longitudinal studies. At the same time, volatile sources should be included for comprehensive coverage and geographical validation by cross-validating source contributions with known infrastructure distributions.

Future research should focus on developing methodologies to balance the trade-off between data diversity and stability. Additional metadata, such as geolocation and latency measurements, could enhance the utility of IPv6 hitlists. Moreover, exploring machine learning techniques to predict IPv6 address stability may improve the hitlist quality. One such approach is called Target Generation Algorithms (TGAs) [2].

References

[1] O. Gasser, J. Zirngibl, and L. Steger, “IPv6 Hitlist Service,” https://ipv6hitlist.github.io/, 2025, [Online; accessed 02-June-2025].

[2] L. Steger, L. Kuang, J. Zirngibl, G. Carle, and O. Gasser, “Target Acquired? Evaluating Target Generation Algorithms for IPv6,” in *Proceedings of the Network Traffic Measurement and Analysis Conference (TMA)*, Jun. 2023, best Paper Award.

[3] J. Gudehege, “Analysis of IPv6 Hitlist sources,” Academic Research Paper, 2024, [Provided by advisor; accessed 02-June-2025].

[4] G. Huston, “Exploring autonomous system numbers,” *The Internet Protocol Journal*, vol. 9, no. 1, pp. 2–23, 2006.

[5] O. Gasser, Q. Scheitle, S. Gebhard, and G. Carle, “Scanning the IPv6 Internet: Towards a Comprehensive Hitlist,” in *Proc. 8th Int. Workshop on Traffic Monitoring and Analysis*, Louvain-la-Neuve, Belgium, Apr. 2016. [Online]. Available: https://net.in.tum.de/pub/ipv6-hitlist/

[6] “Bitnodes ipv6 snapshot dataset,” https://bitnodes.io, 2021, [Online; accessed 13-June-2025].

[7] “Ipinfo.io ipv6 dataset,” https://ipinfo.io/data, 2024, [Online; accessed 13-June-2025].

[8] “Rapid7 forward dns dataset,” https://opendata.rapid7.com/sonar.fdns_v6/, 2022, [Online; accessed 13-June-2025].

[9] “Ripe atlas ipv6 measurement data,” https://atlas.ripe.net, 2024, [Online; accessed 13-June-2025].

[10] “Border Gateway Protocol (BGP),” RFC 1163, Jun. 1990. [Online]. Available: https://www.rfc-editor.org/info/rfc1163

[11] “What is BGP hijacking?” https://www.cloudflare.com/learning/security/glossary/bgp-hijacking/, [Online; accessed 14-June-2025].

[12] K. Z. Sediqi, A. Feldmann, and O. Gasser, “Live long and prosper: Analyzing long-lived moas prefixes in bgp,” 2023, [Online; accessed 05-June-2025].

[13] “Internet health report,” https://www.ihr.live/en/prefix/2001:500:121::/48, [Online; accessed 18-June-2025].

[14] “Internet health report,” https://www.ihr.live/en/prefix/2001:502:f3ff::/48, [Online; accessed 18-June-2025].

[15] “AS2org API Doc,” https://api.data.caida.org/as2org/v1/doc, 2025, [Online; accessed 15-June-2025].

[16] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, “Towards ip geolocation using delay and topology measurements,” in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, 2006, pp. 71–84.