Automated Tracking of the Evolution of Geofeeds: Data Validation and Adoption

Hye Chan Lee and Daniel Park
Department of Computer Science
Calvin University
May 8, 2025

Abstract

Geofeeds are a promising method for enhancing IP geolocation precision by enabling network operators to publish location information for their own IP address ranges. Nevertheless, their use and quality of data are yet to be explored. This paper presents an automated system to test Geofeed data against the standards of RFC 8805 and reports adoption trends based on daily snapshots from CAIDA's WHOIS and routing archives. Our approach incorporates data deduplication, detailed error classification, and longitudinal tracking of Geofeed deployment and data quality.

With our automated validation, we have handled over 1.1 million entries and found 445 erroneous entries in the August 2024 dataset, 91.2% of which were due to invalid subdivision codes. During the period August 2024 to April 2025, the number of unique Geofeed entries increases from 84,712 to 898,204. ASN-level adoption is also boosted, with IPv4 utilization rising from 665 to 7,137 ASNs (more than a tenfold increase), and IPv6 utilization rising from 230 to 1,944 ASNs (nearly an 8.5-fold improvement.

Despite this rapid growth, data quality challenges persist. In April 2025, 98.8% of all validation failures were again attributed to incorrect subdivision codes, indicating that formatting and assignment errors remain the most significant barrier to Geofeed data reliability. Notably, these errors are widespread across a diverse range of network operators, underscoring the need for improved validation tools and clearer guidance to ensure consistent and accurate geolocation data.

1 The IP Geolocation Problem

IP geolocation is the estimation of geographic location based on a machine's IP address. IP geolocation information is widely used in various industries to improve marketing, user experience, and other services. For example, in marketing, geolocation allows businesses to send targeted ads or promotions to users based on their real-time locations [8]. For user experience, IP geolocation enhances navigation and local service recommendations, directing users to nearby options like restaurants, ATMs, or events [28]. In addition to marketing and user experience, IP geolocation has other important use cases. It is used in fraud detection to identify potentially suspicious activities, such as logins or transactions from unexpected locations [16]. E-commerce platforms use geolocation for dynamic pricing and currency conversion, automatically adjusting based on the user's region [19]. Law enforcement and cybersecurity experts utilize geolocation for tracking cyberattacks or locating the source of suspicious traffic [36]. Another significant use case is geoblocking [37], a technology that restricts access to online content based on a user's geographical location, often used to comply with regional copyright laws or licensing agreements.

However, there are significant challenges with IP geolocation accuracy. Due to the use of anycast IP addresses and content delivery networks (CDNs) [29], commercial IP geolocation services often provide imprecise or generalized location data. Anycast inherently routes traffic to the "nearest" or "best-performing" server location, making it impossible to tie an IP address to a single geographic point. This issue is further complicated by the way IP addresses are distributed. Rather than being assigned to specific geographic locations, IP addresses are allocated in blocks to regions and organizations by the Regional Internet Registries (RIRs) [30]. As a result, IPs can be reassigned or rerouted dynamically, making precise geolocation difficult. The use of Network Address Translation (NAT) gateways [31] also introduces additional ambiguity. NAT allows multiple devices on a private network to share a single public IP address, obscuring the true origin of traffic. While NAT conserves IPv4 addresses and enhances security, the shared public IP only reveals the NAT gateway's location—not the individual devices behind it.

Before the introduction of Geofeeds in 2020, the industry relied on three approaches. First, ISPs and network operators provided location hints through WHOIS records and RIR databases, which often became outdated during IP address reassignment [35]. Second, DNS-based approaches, such as EDNS Client Subnet [39], allowed CDNs to approximate user locations, though only for DNS-routed services and with notable privacy trade-offs. Third, active probing [9] attempted to infer geographic positions through latency measurements, but its effectiveness was frequently compromised by anycast routing and multi-path variability.

Geofeeds [25], emerging as a promising solution, enable network operators to self-report IP-to-location mappings in a standardized format. This approach can significantly improve geolocation accuracy for networks with static and well-documented IP assignments. However, Geofeeds are not equipped to resolve the fundamental limitations introduced by anycast and NAT, both of which obscure the true source or destination of traffic. Despite these constraints, Geofeeds may still represent the most practical and scalable alternative to existing geolocation methods, offering greater transparency and accuracy where applicable.

2 What are Geofeeds?

Geofeeds, self-published lists of IP address ranges linked to geolocation details, were introduced by the Internet Engineering Task Force (IETF) [12] in February 2020. They allow network operators to specify the geographic locations of their IP addresses. In other words, rather than inferring IP geolocations based on heuristics or error-prone triangulation methods used by commercial IP geolocation providers (e.g., MaxMind [28], IPinfo [19], and IP2Location [18]), owners of IP addresses share the geographic locations directly. By allowing authoritative data to come from those who manage the IP space, Geofeeds can improve accuracy, reduce ambiguity, and timely provide up-to-date information—particularly in networks with well-documented and relatively static infrastructure.

The current standardized format for Geofeeds is defined in RFC 8805 [22], which specifies that a Geofeed file is a comma-separated values (CSV) file. Each entry contains an IP prefix, along with optional geographic information such as the country code, subdivision code, city, and postal code. Additionally, RFC 9092 [23] introduces a standardized method for publishing Geofeed URLs by registering them with the corresponding Regional Internet Registry (RIR) and including them in WHOIS database entries.

An example of a complete Geofeed entry is:

```
209.112.97.0/24, FR, FR-GES, Strasbourg, 67000
```

Here, `209.112.97.0/24` represents an IP address block in CIDR notation [20]. `FR` is the two-letter country code defined by ISO 3166-1 alpha-2 and `FR-GES` denotes the subdivision code for the Grand Est region as specified by ISO 3166-2, both of which are standardized in ISO 3166 [21]. `Strasbourg` is the city name, and `67000` is the postal code.

The only mandatory field is the IP prefix. Thus, Geofeed entries may still be valid even if the geographic fields are omitted. For example, the IPv6 entries below are syntactically valid because they include proper IP prefixes, even though some geographic fields are missing.

```
2a0c:8840:4000::/34,,,,
2a12:3fc2:ab73::/48,LU,LU-LU,,
```

Furthermore, entries may also be considered invalid if they use non-existent country or subdivision codes, or if the subdivision does not exist within the specified country, as defined by the ISO 3166-2 standard [17].

WHOIS is a collection of regional databases that hold registration details for Internet resources like domain names, IP addresses, and Autonomous System Numbers (ASNs) [34], enabling users to access information about these resources. The database is managed across five regions, each with its own RIR: APNIC (Asia-Pacific region) [4], ARIN (North America) [3], RIPE NCC (Europe) [33], LACNIC (Latin America and the Caribbean) [24], and AFRINIC (Africa) [1]. IANA [13], a department operated by ICANN [11], manages and distributes IPv4 and IPv6 addresses to the regional RIRs. We will use the Geofeed data in the WHOIS database in our research, which we will elaborate on in the following sections.

3 Previous Works on Geofeeds

Geofeeds is a recently introduced standard for publishing IP geolocation data, and academic research on this topic is still in its early stages. Two notable studies contribute to the initial understanding of Geofeed adoption and impact: one by Fainchtein and Sherr [10], and another by Livadariu et al. [25].

Fainchtein and Sherr [10] investigate the adoption of Geofeeds by network operators (ASes) and whether commercial IP-geolocation providers are utilizing them. Their findings indicate significant growth in Geofeed adoption (tenfold over 14 months, from August 2022 to October 2023), but the overall coverage remains limited, including less than 1% of the IPv4 address space. They also observed uneven adoption across regions, with industrialized countries showing higher rates than less industrialized ones. Furthermore, their study reveals a high degree of agreement between Geofeed data and commercial geolocation services (MaxMind-GeoIP2 [28] and IPgeolocation.io [15]), suggesting that these providers may have already incorporated Geofeed data into their databases. They also raise concerns about the lack of cryptographic authentication in publishing Geofeed data, suggesting that the current standards (RFC 8805 [22] and RFC 9092 [23]) may be insufficient for ensuring data integrity.

Livadariu et al. [25] evaluate Geofeeds based on adoption, accuracy, and potential for widespread use, noting that obtaining an Internet-scale accurate IP geolocation dataset has been a longstanding goal. They collected Geofeed data every two weeks over a 3.5-month period—from mid-June to early October 2023—indicating that their analysis overlaps temporally with the dataset used by Fainchtein and Sherr [10]. Their study indicates that Geofeeds are in the early stages of deployment, covering 1.5% of allocated IPv4 prefixes and 0.7% of IPv6 prefixes as of October 2023. These prefixes span 8,316 cities in 249 countries, revealing disparities between continents, with Africa and Latin America showing the smallest fraction of registered Geofeed IPs. They also highlight variability in Geofeed data quality, identifying instances of missing location information and inaccuracies stemming from formatting errors. They suggest that operators share their Geofeeds through WHOIS records, following RFC 9092, to facilitate discovery and improve security. Their analysis reveals that while most Geofeed locations are valid, a fraction of client (0.9%), router (4%), and server (8.5%) IP addresses have incorrect Geofeeds, as determined by comparing them against known locations derived from active measurements. They further suggest that Geofeeds should be considered a supplementary geolocation source rather than a definitive ground truth and provide recommendations for improving the sharing and formatting of Geofeeds.

While both studies acknowledge the potential of Geofeeds for geolocation, they also agree that significant development and refinement are still needed. They identify similar patterns of limited adoption and regional disparities. However, they offer complementary perspectives on data quality and usage. Livadariu et al. focus on internal Geofeed data inconsistencies, while Fainchtein and Sherr examine the relationship between Geofeeds and commercial geolocation services, and the security implications related to RFC 8805 and RFC 9092.

4 Project Contributions

Building on these prior works, our study seeks to address the need for a longer-term analysis by introducing automation to the validation and monitoring of Geofeed data. While the studies in [10] and [25] offer valuable insights into Geofeed adoption, accuracy, and regional disparities, their methodologies present notable limitations. Specifically, both relied on bi-weekly data collection, which may overlook important temporal trends, and lacked automated large-scale validation, limiting the depth and frequency of their assessments. Additionally, it remains unclear whether they employed external tools to map IP addresses to ASNs, which could have enhanced their analyses.

In contrast, our research leverages daily snapshots and the *geofeed-validator* tool [27] to conduct systematic and scalable evaluations of Geofeed entries. By aligning Geofeed records with ASNs [2], we provide a more detailed and dynamic view of Geofeed adoption and evolution over time. Our primary goals are to evaluate Geofeed compliance with RFC 8805, check for duplicate entries, and analyze adoption trends across ASNs using an automated framework. This tool not only categorizes errors by type but also enables a longitudinal perspective on Geofeed accuracy, coverage, and growth. By comparing our findings with those of earlier studies, we highlight shifts in adoption patterns and track improvements or declines in data quality. Ultimately, our approach fills a methodological gap in the literature and introduces new tools and metrics to support future research on the practical viability of Geofeeds.

The objectives of this project are as follows:

- Develop an automated process to validate Geofeed data against RFC 8805 and identify duplicate Geofeed entries.
- Automate the analysis of the adoption of Geofeed data by different ASNs.
- Compare our findings with previous studies in terms of data collection, validation methods, and adoption rates.

5 Data Collection and Methodology

5.1 Geofeed Data Collection

We obtain Geofeed data from the Center for Applied Internet Data Analysis (CAIDA) [6], which collects this data from the WHOIS database. Since October 29, 2024, CAIDA has been using an automated process to pull WHOIS data daily.[1] For our study, we used two datasets: one from August 2024, which was before CAIDA began their automation process, and one from April 7, 2025. The comparisons between these two datasets will be discussed in Section 6 (Comparison of Two Geofeed Snapshots). The August 2024 dataset is referenced here in Section 5 (Data Collection and Methodology). Using the August 2024 dataset alone, we were able to retrieve 652,263 IPv4 entries and 517,200 IPv6 entries. Previous research by Livadariu et al. [25] collected Geofeed data biweekly over 3.5 months using the open-source geofeed-finder tool [26]. Their dataset includes 73,827 IPv4 entries and 46,549 IPv6 entries. Similarly, a study by Fainchtein et al. retrieved Geofeed data biweekly over 14 months using the same tool; however, this study does not specify the number of IPv4 entries and does not provide any IPv6 data.

The first step in our research is to validate the collected data, a step that the two previous studies did not undertake. This validation ensures the accuracy and reliability of the dataset before proceeding with further analysis.

[1] CAIDA classifies Geofeed data as either standard or non-standard based on how RIRs provide Geofeed URLs. If an RIR lacks a dedicated field for Geofeed URLs, the data is classified as non-standard. Since this classification relies on CAIDA's retrieval method, we do not make such distinction in our paper.

5.2 Validation of Geofeed Data

The validation process begins by checking for duplicate entries in the collected Geofeed data. Duplicate entries are identified based on identical IP prefixes and associated location fields. These duplicates can arise from misconfiguration or human error, such as operators unintentionally repeating entries when manually maintaining their Geofeed files. Out of a combined total of 1,169,463 IPv4 and IPv6 Geofeed entries retrieved from the dataset, 81,271 entries contained at least one duplicate. The total number of duplicate occurrences, excluding the first appearance of each entry, was 1,084,751, resulting in 84,712 unique entries. Only these deduplicated entries were passed to the next stage of validation using the open-source *geofeed-validator* tool [27]. This tool checks each Geofeed entry for compliance with RFC 8805, verifies that country and subdivision codes conform to ISO 3166 standards, and ensures that each subdivision code is valid within the specified country.

Previous studies by Livadariu et al. and Fainchtein et al. use the open-source *geofeed-finder* tool [14] to retrieve and validate Geofeed data based on RFC standards. Livadariu et al. identify 135 IPv4 and 143 IPv6 entries lacking location information, which are treated as errors. However, the study by Fainchtein et al. does not report any errors in the dataset. Additionally, neither of these studies addresses the presence of duplicate Geofeed entries in their validation process. In contrast, our research first checks for duplicate entries, retaining only unique Geofeed entries before passing them to the *geofeed-validator* tool. This tool provides more detailed error diagnostics than *geofeed-finder*, validating the unique entries and categorizing errors by type, enabling a more granular analysis of invalid entries.

The *geofeed-validator* checks Geofeed entries against four criteria, returning the following error messages if the validation fails:

- **Not valid subdivision code**: The subdivision code in the Geofeed entry does not conform to the ISO 3166 standards [21]. In ISO 3166, a subdivision code refers to a standardized code that identifies subnational regions (like states, provinces, or territories) within a country.
  - Example: `213.198.107.56/29,RO,RO-RO,Romania,` Error: Not valid Subdivision Code (Here, the subdivision code RO-RO does not exist in ISO 3166)
- **Not valid IP/prefix**: The Geofeed entry contains an invalid IPv4 or IPv6 address range syntax.
  - Example: `102.210.53.128.0/25,ZA,ZA-GP,Johannesburg,` Error: Not valid IP/prefix (Here, the IPv4 address has an invalid syntax)
- **The subdivision is not inside the country**: The provided subdivision code does not match the country code.
  - Example: `2404:f4c0:fd06::/48,TW,CN-TW,Taipei,` Error: The Subdivision is not inside the Country (Here, the subdivision CN-TW does not exist in the country code TW)
- **Not valid country code**: The country code in the Geofeed entry is not valid per ISO 3166.
  - Example: `2a0d:1a40:7800:40::/58,EU,,,` Error: Not valid Country Code (Here, the country code EU is invalid)

From the 84,712 unique entries in the August 2024 Geofeed dataset obtained from CAIDA, we validated 84,267 as correct using the *geofeed-validator* tool—60,347 IPv4 entries and 23,920 IPv6 entries. The tool identified 445 errors, including 275 IPv4 entries and 170 IPv6 entries, which we excluded from further analysis. We use this deduplicated August 2024 dataset as the foundation for our validation and research because it was collected before CAIDA began automating Geofeed data retrieval from the WHOIS database. All statistics and methodology in our Data Collection and Methodology section build on this data.

5.3 Error Analysis

Unlike previous studies, our research focuses more deeply on categorizing and analyzing invalid Geofeed data. We not only examine the sources and causes of invalidation but also provide insights into why specific Geofeed data do not meet the required standards, contributing to a better understanding of Geofeed data quality.

Table 1 shows the percentage distribution of different types of errors returned after running the *geofeed-validator* on deduplicated Geofeed entries. As shown in the table, more than 90% of the invalidated Geofeed data result from invalid subdivision codes. The table combines error counts from both IPv4 and IPv6 Geofeed entries. Because Geofeed files often contain a mixture of IPv4 and IPv6 addresses, and we process the files obtained from CAIDA as a whole through the *geofeed-validator*, we do not separate errors by the address family.

We quantify the occurrence of each error type. The breakdown of errors is as follows:

Error Type	Number of Occurrences	Percentage
Not valid subdivision code	406	91.2%
Not valid IP/prefix	6	1.4%
The subdivision is not inside the country	5	1.1%
Not valid country code	28	6.3%

Table 1: Counts and percentages of error types for combined IPv4 and IPv6 Geofeed data.

Since Geofeed data consists of self-published IP-geolocation information, it is generally assumed to be accurate. This assumption likely explains why previous studies have not addressed the potential for human error. However, during our analysis, we identify a Geofeed entry with an IPv4 prefix of /2—an unlikely and notable anomaly, as shown in Figure 2. A /2 prefix represents 2^{30} IP addresses, equivalent to a quarter of the entire IPv4 address space. It is highly improbable for a single organization to control such a large range, strongly suggesting an error in the prefix length.

This discovery implies that additional human errors may exist within the dataset. Nevertheless, aside from this case, we do not identify any other significant anomalies. To ensure the integrity of our analysis, we exclude the /2 entry and apply additional filters: specifically, we remove any Geofeed entries with IPv4 prefixes shorter than /8 and IPv6 prefixes shorter than /19. These thresholds are based on typical allocation sizes by Regional Internet Registries (RIRs) [30], where /8 and /19 represent the largest prefixes typically delegated to users. As a result, only one IPv4 entry is excluded, and the remaining data is incorporated into our analysis.

As a result, Figure 1 shows that the August 2024 snapshot provides a total of 60,346 valid IPv4 entries—representing 9.3% of the original IPv4 dataset—and 23,920 valid IPv6 entries—representing 4.6% of the original IPv6 dataset. These percentages indicate the proportion of valid entries relative to the original dataset.

Original Dataset
IPv4: 652,263   IPv6: 517,200

↓

After excluding duplicate entries
IPv4: 60,622 (9.3%)   IPv6: 24,090 (4.6%)

↓

After running `geofeed-validator`
IPv4: 60,347 (9.3%)   IPv6: 23,920 (4.6%)

↓

After removing prefix length errors
IPv4: 60,346 (9.3%)   IPv6: 23,920 (4.6%)

Figure 1: IPv4 and IPv6 Geofeed entry counts after validation.

5.4 Adoption of Geofeeds by IP Addresses

We perform a similar analysis as from the research done by Livadariu et al. [25] to determine the count and percentage of IPv4 and IPv6 address ranges. Tables 2 and 3 are based on Geofeed entries that comply with RFC 8805 standards.

Among the Geofeed data validated by the *geofeed-validator*, we analyze the number and percentage of IPv4 and IPv6 address ranges, as summarized in Tables 2 and 3. For IPv4, the most common prefix lengths are /24 (17.0%), /31 (15.9%), and /32 (37.6%). For IPv6, the most common prefix lengths are /44 (9.8%), /56 (36.4%), and /64 (33.5%). The distribution of prefix lengths reflects the common operational practices. In IPv4, /24 is widely used because it is the smallest subnet size that allows for routing between networks, providing enough host addresses for small to medium-sized networks. /31 is commonly used for point-to-point links, providing just two usable addresses, while /32 is often used for individual IP addresses, particularly in routing or network address translation (NAT) scenarios. For IPv6, /56 is frequently assigned to end users or organizations as it provides a sufficiently large address space for most needs, while /44 is commonly used in larger-scale deployments. Finally, /64 is the standard subnet size, as recommended by IPv6 addressing architecture, and is used for most subnets to allow for efficient routing and address assignment. In accordance with RFC 8805, our analysis also includes Geofeed entries that specify a single IP address without an explicit prefix length, treating them as /32 for IPv4 or /128 for IPv6 in our statistics.

IPv4 Prefix	/2	/11	/15	/16	/17	/18	/19	/20	/21	/22	/23	/24	/25	/26	/27	/28	/29	/30	/31	/32
Count	1	1	3	38	69	147	313	509	825	1,359	1,386	8,679	702	1,085	1,707	1,685	2,942	2,332	8,104	19,174
Percentage	0.002	0.002	0.006	0.074	0.135	0.288	0.613	0.997	1.616	2.662	2.714	16.997	1.375	2.125	3.343	3.300	5.762	4.567	15.871	37.551

Table 2: IPv4 prefix distribution for valid entries.

To compare our findings with the prior works, we reference the results reported by Livadariu et al. For IPv4, both Livadariu's study and ours observe /24 and /32 as the most prevalent prefix lengths, with similar proportions. Livadariu et al. report that /24 prefixes account for 20.5% and /32 prefixes for 33.4%, closely matching our observations of 17.0% and 37.6%, respectively. However, while we identify /31 (15.9%) as the second most common prefix length, Livadariu et al. report /29 (14.9%) instead. For IPv6, both studies find

IPv6 Prefix	/28	/29	/30	/31	/32	/33	/34	/35	/36	/37	/38	/39	/40	/41	/42	/43	/44	/45	/46	/47
Count	3	121	2	3	244	11	34	25	109	23	46	23	236	3	3	84	2,341	370	56	434
Percentage	0.013	0.506	0.008	0.013	1.020	0.046	0.142	0.105	0.456	0.096	0.192	0.096	0.987	0.013	0.013	0.351	9.787	1.547	0.234	1.814

IPv6 Prefix	/48	/49	/50	/51	/52	/53	/54	/55	/56	/57	/58	/59	/60	/63	/64	/96	/125	/126	/127	/128
Count	2,145	72	3	8	27	9	10	391	8,706	1	4	27	142	2	8,013	1	3	11	58	116
Percentage	8.967	0.301	0.013	0.033	0.113	0.038	0.042	1.635	36.396	0.004	0.017	0.113	0.594	0.008	33.499	0.004	0.013	0.046	0.242	0.485

Table 3: IPv6 prefix distribution for valid entries.

/56 and /64 prefixes to be dominant. Livadariu et al. report /56 prefixes comprising 48.2% and /64 prefixes 33.4% of the dataset, whereas we observe 36.4% for /56 and 33.5% for /64. Additionally, Livadariu et al. observe /48 (6.9%) as a notable prefix length, while our analysis identifies /44 (9.8%) more prominently. These differences may reflect variations in dataset composition, Geofeed publishing practices, and address allocation trends over time.

5.4.1 Adoption of Geofeeds by ASNs

To map IPv4 adn IPv6 address ranges in Geofeed data to their respective ASNs, we utilize CAIDA's pfx2as tool [7], which provides a daily updated list of address prefixes and their associated ASNs. This tool enables us to accurately and efficiently associate each geofeed entry with the ASN responsible for announcing the corresponding IP address space on the Internet.

We identify the ASNs for both IPv4 and IPv6 address ranges—60,621 IPv4 entries and 24,090 IPv6 entries-by combining valid and invalid Geofeed data and mapping them using the pfx2as tool. We include both valid and invalid Geofeed entries in ASN identification as shown in step 2 of Figure 2, as our goal is to analyze the overall Geofeed coverage per AS and determine the proportion of invalid Geofeed data within ASes that publish Geofeed information. To avoid double-counting in cases where Geofeed entries contain overlapping IP prefixes, we retain only the entry with the longest prefix length. For instance, in the August 2024 snapshot, ASN `202265` had three matching entries: `94.199.1.0/24`, `94.199.1.226/32`, and `94.199.1.225/32`. Since the /32 prefixes are subsets of the /24, only the /24 is included in the analysis.

From the 445 Geofeed entries that fail to pass the *geofeed-validator* (275 IPv4 and 170 IPv6), we extract only the IP prefix field. We then check the syntax of the IP prefixes. All 445 entries had correct IP prefix syntax, so we were able to retain all 275 IPv4 and 170 IPv6 entries. We then merge these entries with the validated Geofeed data, resulting in a total of 60,621 IPv4 entries and 24,090 IPv6 entries. This combined dataset serves as the basis for our ASN mapping and analysis.

The previous study by Livadariu et al. extracts information from Regional Internet Registry (RIR) delegation data [32] to map IP addresses in the Geofeed data to Autonomous Systems (ASes). Similarly, Fainchtein et al. use the Stanford ASdb [38] to classify the ASes associated with Geofeed data. In contrast, we utilize CAIDA's pfx2as tool [7] to directly map IP address ranges from the Geofeed data to ASNs. This approach enables us to evaluate Geofeed data coverage across different ASNs while leveraging the most recently updated and operationally relevant IP-to-ASN mappings. Because pfx2as derives its mappings from real-time BGP routing table [5] snapshots, it provides a more accurate and current reflection of how IP address space is used and announced on the Internet, compared to static registry-based methods or databases that may lag behind actual routing behavior. Additionally, as CAIDA pulls WHOIS data and updates ASN mapping information daily, our automation ensures that the most recent datasets are always utilized in our analysis.

During the ASN identification, we encounter three unexpected cases:

- **Case 1**: CAIDA's pfx2as tool does not return a matching ASN for some IPv4 and IPv6 addresses, likely because these IP address ranges are not routable and are therefore absent from the Border Gateway Protocol (BGP) routing tables [5]. For example, the IPv4 address range `154.83.190.0/24` and the IPv6 address range `2a0f:7802:e123::/48` are among those that lack a matching ASN. Consequently, we exclude these addresses from our analysis, as shown in step 2 of Figure 2 ("Remove no ASN match"). Specifically, 676 out of 60,621 IPv4 entries (1.1%) and 1,832 out of 24,090 IPv6 entries (7.6%) are excluded due to the absence of a matching ASN.

- **Case 2**: CAIDA's pfx2as tool returns multiple ASNs for some IPv4 addresses. For example, the IP address range `209.145.44.0/22` corresponds to the following ASNs: `202656`, `29802`, `21859`, and `136557`. This occurs because the pfx2as tool maps IP addresses to ASNs at the subnet level. However, in cases where a network provider's prefix encompasses IP ranges from multiple ASNs, including those of its customers, the tool may return multiple ASNs for a single address, as it cannot determine which ASN should be assigned to that specific address. This limitation is inherent to the tool's design and cannot be addressed within its current functionality. To prevent potential statistical errors, as depicted in step 3 of Figure 2 ("Remove multiple ASN match"), we exclude these addresses from our analysis, which consists of 38 IPv4 addresses. Additionally, no overlapping ASNs are found for IPv6 addresses.

- **Case 3**: Some Geofeed entries have a shorter prefix than the IP prefix allocated to an AS, which causes the Geofeed entry to cover more IP addresses than the AS actually owns. For example, ASN `3970` owns 256 IP addresses with a /24 prefix. However, the IP address matching this ASN is `45.132.188.0/22`, which has a shorter prefix, thus covering more IP addresses. The reason this IP address matched only one ASN was that there were no other ASes covering the remaining IP addresses within that prefix length. To ensure our analysis reflects actual ownership, we excluded these ASes from our study as shown in step 4 of Figure 2 ("Remove shorter prefixes than matching ASN"). As a result, we removed 15 ASes and a total of 84 IPv4 Geofeed entries. This issue was only observed in IPv4 and no such cases were found for IPv6.

As shown in Figure 2, we analyze the remaining 59,823 IPv4 and 22,258 IPv6 entries to evaluate the overall Geofeed coverage per AS, matching 665 ASNs with IPv4 addresses and 230 ASNs with IPv6 addresses.

Invalid Geofeed Entries
IPv4: 275 IPv6: 170

↓ Step 1: Add to validated entries count

After adding to the validated Geofeed entries
IPv4: 60,622 (60,146 valid + 275 invalid)
IPv6: 24,090 (23,920 valid + 170 invalid)

↓ Step 2: Remove no ASN match

After removing IP addresses with no ASN match
IPv4: 59,945 IPv6: 22,258

↓ Step 3: Remove multiple ASN match

After removing IP addresses with multiple ASN match
IPv4: 59,907 IPv6: 22,258

↓ Step 4: Remove shorter prefixes than matching ASN

After removing Geofeed entries with shorter prefixes than their matching ASNs
IPv4: 59,823 IPv6: 22,258

↓ Step 5: Show ASN match count

ASN Match Counts for Remaining Addresses
IPv4: 665 ASNs matched
IPv6: 230 ASNs matched

Figure 2: Summary of filtering Geofeed entries and matching ASNs.

5.4.2 What percentage of IP addresses within each ASN are covered by Geofeed data?

Table 4 shows that 0.9% of the ASes provide at least one IPv4 Geofeed data entry that covers a portion of their IPv4 address range. The distribution of ASNs by IPv4 Geofeed coverage percentage is illustrated in Figure 3. The x-axis represents the Geofeed coverage percentage, indicating the proportion of IP addresses owned by an AS that are covered in the Geofeed data, while the y-axis shows the number of ASes that fall within each coverage range on a logarithmic scale. Notably, 221 ASNs provide Geofeed data for less than 10% of the IPv4 address range they manage, accounting for about one-third of all ASNs sharing Geofeed data. Interestingly, the second highest count is observed in the 90–100% coverage range, where 159 ASNs provide near-complete Geofeed data for their IPv4 space.

	AS Count	Percentage
Has Geofeed Coverage	665	0.9%
No Geofeed Coverage	76,086	99.1%

Table 4: Count and percentages of ASNs that have at least one IPv4 Geofeed entry.

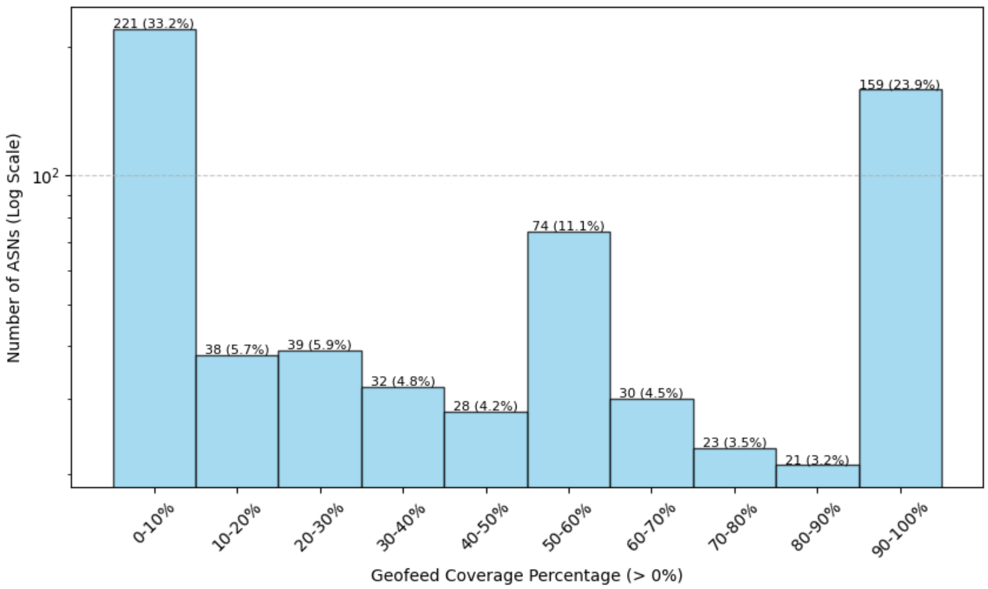

Figure 3: Distribution of ASNs by the percentageof IPv4 Geofeed coverage.

Table 5 indicates that only 0.7% of ASes contribute at least one IPv6 Geofeed entry covering part of their IPv6 address range. Figure 4 illustrates a scatterplot showing the distribution of IPv6 Geofeed coverage across ASNs, grouped by ASN size. The x-axis represents the log base 2 of the number of Geofeeds in /64 IPv6 addresses, while the y-axis shows the log base 10 count of ASNs that provide Geofeed data, with each point's color indicating the size of the ASN based on the number of /48 prefixes it advertises. This visualization helps to explore the correlation between ASN size and the extent of Geofeed coverage.

From the chart, it is clear that ASNs with larger address allocations tend to cover more IPv4 addresses. Specifically, ASNs with between 10k-100k and 100k-1M /48 prefix counts provide the most IPv4 addresses in their Geofeed data. In contrast, ASNs with fewer than 10k /48 prefixes do not show a noticeable trend in the number of IPv4 addresses covered. Additionally, the chart highlights that many ASes provide Geofeed data in the range between 15 and 20 on the x-axis, which corresponds to the logarithmic scale of Geofeed coverage in /64 IPv6 prefixes.

	Coverage Count	Percentage
Has Geofeed Coverage	230	0.7%
No Geofeed Coverage	34,442	99.3%

Table 5: Count and percentages of ASNs that have at least one IPv6 Geofeed entry.

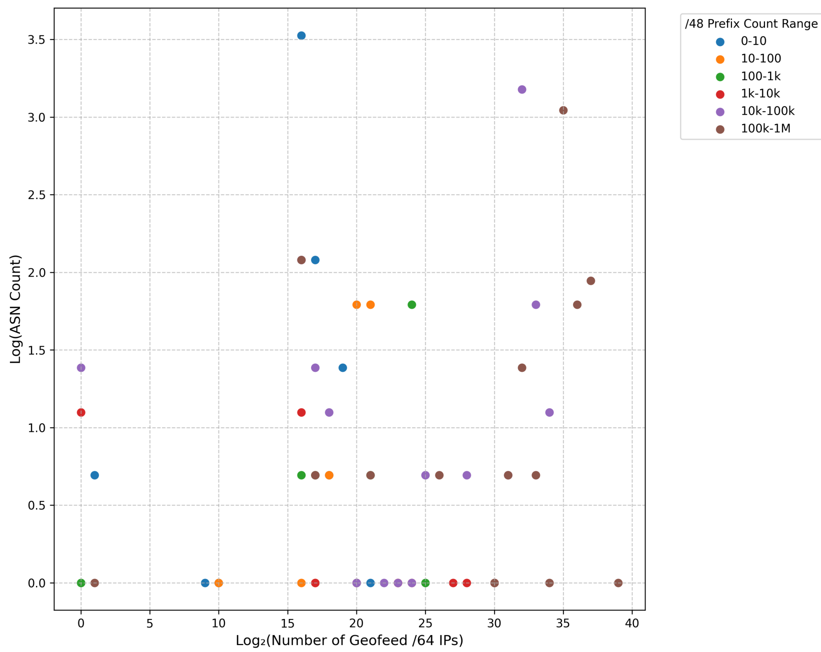

Figure 4: Distribution of ASNs by IPv6 Geofeed coverage percentage.

5.4.3 What percentage of the Geofeed data within each AS fails validation checks in *geofeed-validator*?

Table 6 shows that 7.4% of the ASNs that provided Geofeed data include at least one invalid IPv4 Geofeed entry. For those ASNs, the error distribution of the Geofeed data is shown in Figure 5. Notably, 20 ASNs provided Geofeed data with invalid IPv4 entries comprising 90% to 100% of their submissions. This high error rate is primarily due to these ASNs not adhering to the RFC8805 standardized format for publishing Geofeed entries, resulting in widespread validation failures. Interestingly, the second highest bar in the figure corresponds to the 0–10% error range, with 19 ASNs. This pattern suggests a polarization among ASNs with invalid data: while those in the 90–100% range are likely failing to follow RFC8805 entirely, those in the 0–10% range have mostly valid Geofeed entries but include a small number of entries with syntax or formatting errors.

	AS Count	Percentage
Has No Invalid Data	616	92.6%
Has at least 1 Invalid Data	49	7.4%

Table 6: Count and percentages of ASNs providing at least one IPv4 Geofeed entry.

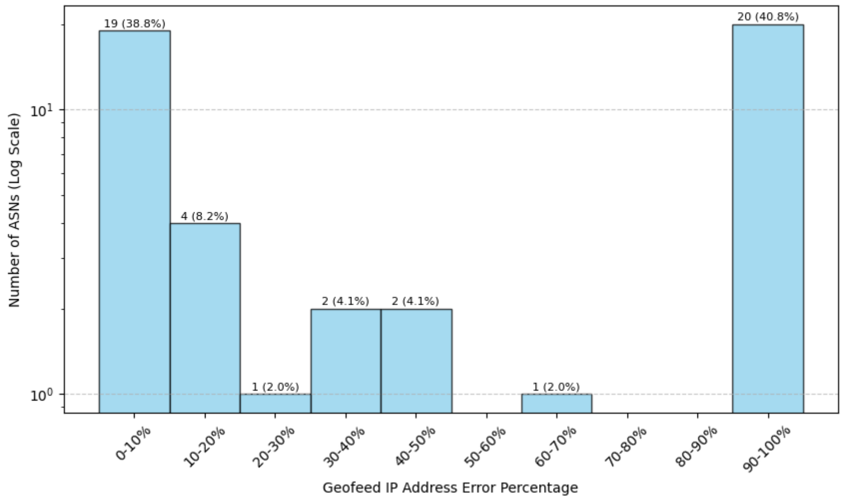

Figure 5: Distribution of errors in the ASNs that has invalid IPv4 Geofeed data.

Table 7 indicates that 14.8% of ASNs supplying Geofeed data contain invalid IPv6 entries. The error distribution for these ASNs is illustrated in Figure 6. Notably, 16 ASNs reported Geofeed data in which 90% to 100% of the IPv6 entries were invalid, suggesting a failure to comply with the RFC8805 specification. The second largest group falls within the 0–10% error range, comprising 11 ASNs. This polarization—where most errors are concentrated either in near-complete invalidity or near-complete validity—is consistent with the IPv4 findings.

	AS Count	Percentage
Has No Invalid Data	196	85.2%
Has at least 1 Invalid Data	34	14.8%

Table 7: Count and percentages of ASNs providing at least one IPv6 Geofeed entry.

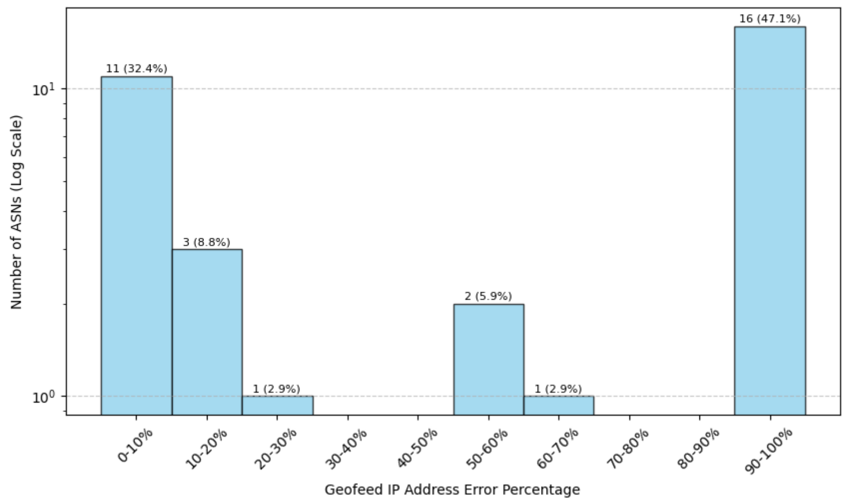

Figure 6: Distribution of errors in the ASNs that has invalid IPv6 Geofeed data.

6 Comparison of Two Geofeed Snapshots

6.1 Geofeed Data Retrieval Comparison

We conducted a comparison of the total number of IPv4 and IPv6 Geofeed entries obtained, collected on August 2024 (used in the previous section) and April 7, 2025. The purpose of this comparison is to analyze the change in Geofeed adoption over an 8-month period. The breakdown of the data is as follows:

Table 8 shows that over the span of eight months, approximately 320,000 additional Geofeed entries are retrieved, indicating substantial growth in the dataset. Both IPv4 and IPv6 entries show notable increases, with IPv4 entries rising from 652,263 to 902,100 and IPv6 entries from 517,200 to 593,410, suggesting broader address space coverage. Notably, the number of unique Geofeed entries increases by an order of magnitude—from 84,712 in August 2024 to 898,204 by April 2025. This increase is reflected in the growth of unique IPv4 and IPv6 Geofeed entries, which rise from 60,622 to 544,766 and from 24,090 to 353,438, respectively. These developments underscore not only the expanding adoption of Geofeeds across a wider range of network operators but also improvements in the granularity and comprehensiveness of the collected data.

	Snapshot from August 2024	Snapshot from April 7, 2025
Total Number of Geofeed Entries	1,169,463	1,495,510
Total IPv4 Geofeed Entries	652,263 (55.8%)	902,100 (60.3%)
Total IPv6 Geofeed Entries	517,200 (44.2%)	593,410 (39.7%)
Total Unique Geofeed Entries	84,712	898,204
Total Unique IPv4 Geofeed Entries	60,622 (71.6%)	544,766 (60.7%)
Total Unique IPv6 Geofeed Entries	24,090 (28.4%)	353,438 (39.3%)

Table 8: Comparison of Geofeed data entries retrieved for August 2024 and April 7, 2025.

6.2 Error Analysis Comparison

Compared to the August 2024 snapshot, the April 7, 2025 snapshot reports 98,049 errors out of 898,204 unique entries, resulting in a higher error rate of approximately 10.9%. Despite the rise in validation errors, the number of valid entries increases substantially: valid IPv4 entries grow from 60,347 to 451,100 (a 7.5-fold increase), and valid IPv6 entries increase from 23,920 to 349,055 (a more than 14-fold increase).

While the total number of errors rises sharply, the distribution of error types remains heavily concentrated in a single category. As shown in Table 9, the majority of validation issues in both snapshots stem from the not valid subdivision code" error, which accounts for 91.2% of errors in August 2024 and increases to 98.8% by April 2025. Other error types remain relatively rare: not valid country code" errors increase from 28 to 531, and not valid IP/prefix" errors rise from 6 to 374. The number of subdivision is not inside the country" errors increases more significantly—from 5 to 253—but still comprises only a small fraction of the total. This indicates that as Geofeed data scales, data quality—particularly the correct specification of subdivisions—emerges as a key area for improvement. One potential way to reduce the number of such errors is to update the *geofeed-validator* tool to use the most recent ISO 3166 information when checking subdivision and country codes.

	Snapshot from August 2024		Snapshot from April 7, 2025	
Error Type	Number of Occurrences	Percentage	Number of Occurrences	Percentage
Not valid subdivision code	406	91.2%	96,891	98.8%
Not valid IP/prefix	6	1.4%	374	0.4%
The subdivision is not inside the country	5	0.3%	253	3.4%
Not valid country code	28	6.3%	531	0.5%

Table 9: Comparison between count and percentages of error types in Geofeed data between August 2024 and April 7, 2025.

6.3 IP Address Prefix Distribution Comparison

Tables 2 and 10 display the IPv4 prefix distribution for the August 2024 and April 7, 2025 snapshots, respectively. The April 2025 snapshot (Table 10) identifies /32 (24.6%), /29 (24.4%), and /24 (17.0%) as the most prevalent prefix lengths. In contrast, the August 2024 data are dominated by /32 (37.6%), followed by /24 (17.0%) and /31 (15.9%). This shift indicates a notable decline in the proportion of /32 prefixes and a rise in /29 usage, suggesting a move toward specifying slightly broader address blocks. The counts for IPv4 prefixes all increase in the April 2025 snapshot, with the greatest increase occurring in the /32 prefix. The count for /32 prefixes rises from 19,174 in August 2024 to 102,819 in April 2025. This substantial increase in the /32 prefix count further reflects the growing use of these smaller address blocks in Geofeed data. Also, unlike the August snapshot, the April 2025 data do not exhibit any apparent errors in prefix length, suggesting improved data quality. These April results are more consistent with the findings of Livadariu et al., who report /32 and /24 as dominant IPv4 prefixes, comprising 33.4% and 20.5% of entries, respectively, and also highlight /29 (14.9%) as a frequent prefix length. The alignment between the April 2025 data and prior research highlights similar trends in Geofeed prefix usage and allocation.

IPv4 Prefix	/2	/11	/15	/16	/17	/18	/19	/20	/21	/22	/23	/24	/25	/26	/27	/28	/29	/30	/31	/32
Count	1	1	3	38	69	147	313	509	825	1,359	1,386	8,679	702	1,085	1,707	1,685	2,942	2,332	8,104	19,174
Percentage	0.002	0.002	0.006	0.074	0.135	0.288	0.613	0.997	1.616	2.662	2.714	16.997	1.375	2.125	3.343	3.300	5.762	4.567	15.871	37.551

Table 2: IPv4 prefix distribution from August 2024 snapshot.

IPv4 Prefix	/9	/10	/11	/12	/13	/14	/15	/16	/17	/18	/19	/20	/21	/22	/23	/24	/25	/26	/27	/28	/29	/30	/31	/32
Count	1	1	1	6	3	12	41	307	292	549	1,415	2,370	4,257	7,963	9,665	73,550	5,407	7,820	13,431	23,963	102,118	42,828	18,981	102,819
Percentage	0.000	0.000	0.000	0.001	0.001	0.003	0.010	0.073	0.070	0.131	0.339	0.567	1.019	1.906	2.313	17.604	1.294	1.872	3.215	5.736	24.442	10.251	4.543	24.610

Table 10: IPv4 prefix distribution from April 7, 2025 snapshot.

Tables 3 and 11 display the IPv6 prefix distribution for the August 2024 and April 7, 2025 snapshots, respectively. Compared to the August 2024 snapshot, the April 2025 snapshot (Table 11) shows a shift, with /63 (13.2%), /64 (35.3%), and /112 (10.8%) as the leading prefixes. Despite these changes, the /64 prefix remains dominant in both snapshots. The counts for IPv6 prefixes also increase in the April 2025 snapshot, with the most notable rise occurring in the /64 prefix length. In August 2024, the count for /64 was 8,013, while in April 2025, it increased to 123,303. This substantial increase further highlights the growing prevalence of the /64 prefix. The distribution observed in the August 2024 snapshot is more consistent with the findings of Livadariu et al., who report /56 and /64 as the most common prefixes. In their study, /56 accounts for 48.2% and /64 for 33.4%, which closely aligns with our August 2024 results. However, the April 2025 snapshot introduces new trends with the emergence of /63 and /112 prefixes, which were not noted in prior research. These differences may reflect changes in Geofeed publishing practices, address allocation trends, or variations in dataset composition.

IPv6 Prefix	/28	/29	/30	/31	/32	/33	/34	/35	/36	/37	/38	/39	/40	/41	/42	/43	/44	/45	/46	/47
Count	3	121	2	3	244	11	34	25	109	23	46	23	236	3	3	84	2,341	370	56	434
Percentage	0.013	0.506	0.008	0.013	1.020	0.046	0.142	0.105	0.456	0.096	0.192	0.096	0.987	0.013	0.013	0.351	9.787	1.547	0.234	1.814

IPv6 Prefix	/48	/49	/50	/51	/52	/53	/54	/55	/56	/57	/58	/59	/60	/63	/64	/96	/125	/126	/127	/128
Count	2,145	72	3	8	27	9	10	391	8,706	1	4	27	142	2	8,013	1	3	11	58	116
Percentage	8.967	0.301	0.013	0.033	0.113	0.038	0.042	1.635	36.396	0.004	0.017	0.113	0.594	0.008	33.499	0.004	0.013	0.046	0.242	0.485

Table 3: IPv6 prefix distribution from August 2024 snapshot.

IPv6 Prefix	/28	/29	/30	/31	/32	/33	/34	/35	/36	/37	/38	/39	/40	/41	/42	/43	/44	/45	/46	/47	/48	/49	/50	/51	/52	/53	/54	/55
Count	6	459	49	27	2,679	34	85	123	1,198	327	184	259	3,185	125	260	154	6,672	35,061	657	851	15,377	224	77	11	8,117	73	98	734
Percentage	0.002	0.132	0.014	0.008	0.768	0.010	0.024	0.035	0.343	0.094	0.053	0.074	0.913	0.036	0.075	0.044	1.912	10.048	0.188	0.244	4.407	0.064	0.022	0.003	2.326	0.021	0.028	0.210

IPv6 Prefix	/56	/57	/58	/59	/60	/61	/62	/63	/64	/68	/76	/79	/80	/96	/110	/112	/113	/116	/118	/119	/120	/122	/123	/124	/125	/126	/127	/128
Count	10,679	3	5	332	5,089	10,321	25,547	46,165	123,303	1	2	1	7	55	59	37,846	6	1	3	8	94	3	3	35	89	1,363	7,201	3,621
Percentage	3.060	0.001	0.001	0.095	1.458	2.958	7.321	13.230	35.336	0.000	0.001	0.000	0.002	0.016	0.017	10.846	0.002	0.000	0.001	0.002	0.027	0.001	0.001	0.010	0.026	0.391	2.064	1.038

Table 11: IPv6 prefix distribution from April 7, 2025 snapshot.

6.4 ASN Analysis Comparison

The ASN analysis includes both valid and non-valid Geofeed data. For the August 2024 snapshot, we used a total of 60,622 IPv4 and 24,090 IPv6 entries, while for the April 7, 2025 snapshot, we used 544,766 IPv4 and 353,438 IPv6 entries.

Three specific cases led to the exclusion of certain IPv4 and IPv6 entries from our ASN analysis as shown in Table 12. First, some IP addresses did not return any matching ASN when processed using CAIDA's pfx2as tool. This typically occurs when the IP ranges are not present in BGP routing tables and are therefore considered unrouted. In the August 2024 snapshot, 676 IPv4 and 1,832 IPv6 entries were excluded for this reason. By April 7, 2025, this number increased significantly to 4,302 IPv4 and 8,199 IPv6 entries. Second, a small number of IPv4 addresses returned multiple ASN matches, indicating ambiguous mappings caused by overlapping subnet allocations. Specifically, 38 IPv4 entries were excluded in August 2024, and 253 were excluded in April 2025. No such overlaps were observed for IPv6 addresses in either snapshot. Third, some IPv4 Geofeed entries had shorter prefix lengths than the prefixes officially associated with the matched ASNs, meaning these entries potentially referenced a broader IP range than the ASN was actually assigned. This mismatch led to the exclusion of 84 IPv4 entries in August 2024 and 2,747 in April 2025. No such cases were found in IPv6 data. In total, 798 IPv4 (1.3%) and 1,832 IPv6 (7.6%) entries were excluded in August 2024, and 7,302 (1.3%) IPv4 and 8,199 IPv6 (2.3%) entries were excluded in the April 2025 snapshot.

Case	Description	Aug 2024	Apr 7, 2025
Case 1	No matching ASN (IPv4/IPv6)	676 / 1,832	4,302 / 8,199
Case 2	Multiple matching ASNs (IPv4 only)	38	253
Case 3	Prefix length shorter than ASN allocation (IPv4 only)	84	2,747
	Total excluded IPv4 entries (% of total)	**798 (1.3%)**	**7,302 (1.3%)**
	Total excluded IPv6 entries (% of total)	**1,832 (7.6%)**	**8,199 (2.3%)**

Table 12: Number of excluded entries by case and snapshot date.

As a result, 59,824 IPv4 and 22,258 IPv6 entries from the August 2024 snapshot, and 537,464 IPv4 and 345,239 IPv6 entries from the April 7, 2025 snapshot were retained for the subsequent ASN analysis.

6.4.1 ASN Geofeed Coverage Percentage Comparison

Table 13 compares the IPv4 Geofeed coverage among Autonomous Systems between the August 2024 and April 7, 2025 snapshots. The number of ASes with Geofeed coverage increases significantly from 665 (0.9%) in August 2024 to 7,137 (9.3%) in April 2025. This more than tenfold growth indicates a substantial improvement in Geofeed adoption. However, the table also shows that the majority of ASes—over 90% in both snapshots—still lack IPv4 Geofeed coverage, highlighting the continued need for broader participation and more consistent publishing of Geofeed data.

Coverage Type	Snapshot from August 2024		Snapshot from April 7, 2025	
	Count	Percentage	Count	Percentage
Has Geofeed Coverage	665	0.9%	7,137	9.3%
No Geofeed Coverage	76,086	99.1%	69,418	90.7%

Table 13: Comparison of ASNs with and without IPv4 Geofeed coverage between August 2024 and April 7, 2025.

Figures 3 and 7 show the distribution of Autonomous Systems (ASNs) by their IPv4 Geofeed coverage percentage, comparing snapshots from August 2024 and April 7, 2025. The distributions both have a bi-modal shape with many ASNs clustering at either minimal (0-10%) or near-complete (90-100%) coverage. Notably, most changes between the two snapshots occurr at these extremes.

In August 2024 (Figure 3), 33.2% of ASNs with Geofeed data report less than 10% coverage, while 23.9% achieve near-complete (90–100%) coverage. The intermediate ranges remain relatively less prevalent. By April 2025 (Figure 7), ASNs with near-complete coverage rises to 38.4%, while those with minimal coverage drop to 14.3%. This shift indicates that many ASNs actively enhance their Geofeed completeness or adopt high-coverage entries, indicating an overall improvement in the coverage.

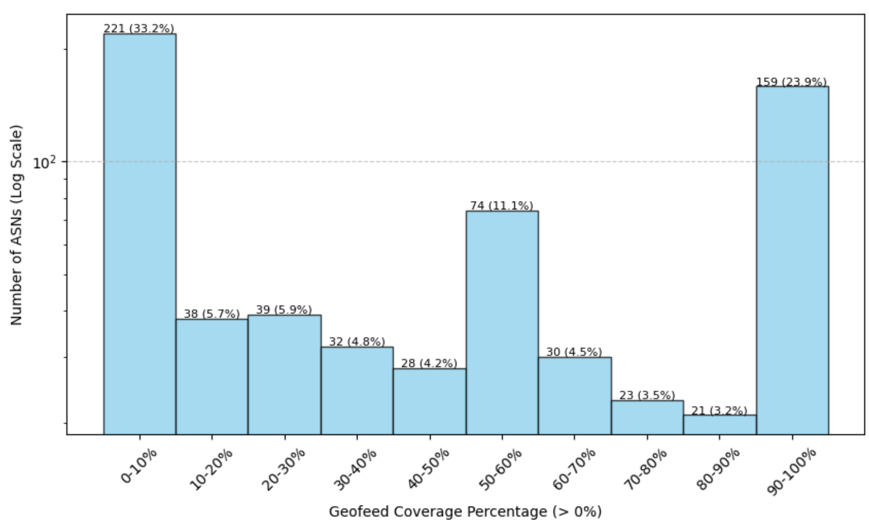

Figure 3: Distribution of ASNs by IPv4 Geofeed coverage percentage from August 2024 snapshot.

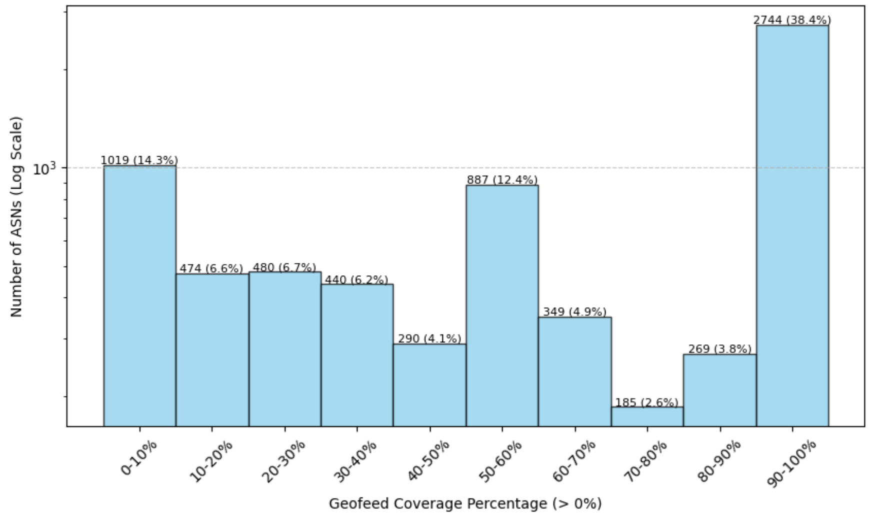

Figure 7: Distribution of ASNs by IPv4 Geofeed coverage percentage from April 7, 2025 snapshot.

As shown in Table 14, the number of ASNs providing IPv6 Geofeed coverage increases from 230 (0.7%) in August 2024 to 1,944 (5.6%) by April 7, 2025—a twelvefold growth. Between the periods of the two snapshots, the dataset saw an increase of approximately 500,000 IPv6 entries, which contributes to the rise in ASN adoption. Despite this improvement, the majority of ASNs—over 94%—still do not include IPv6 Geofeed data.

Coverage Type	Snapshot from August 2024		Snapshot from April 7, 2025	
	Coverage Count	Percentage	Coverage Count	Percentage
Has Geofeed Coverage	230	0.7%	1,944	5.6%
No Geofeed Coverage	34,442	99.3%	32,728	94.4%

Table 14: Comparison of ASN coverage for IPv6 Geofeed data between August 2024 and April 7, 2025.

As shown by Figures 4 and 8, the snapshot from April 7, 2025, reveals a significant increase in the number of IPv6 Geofeed entries compared to the August 2024 data, indicating broader adoption and coverage of IPv6 addresses. The 100k-1M /48 prefix range remains dominant, continuing to show the highest Geofeed coverage. However, unlike the August 2024 snapshot, the ASNs with fewer than 100k /48 prefixes do not exhibit a clear trend in coverage. The April 2025 data also highlights a concentration of Geofeed data in the 15-20 range on the x-axis, corresponding to the log scale of Geofeed coverage in /64 IPv6 prefixes. While this trend also appears in the August 2024 chart, the April 2025 data shows a significantly greater density in this range. Overall, the April 7, 2025 snapshot reflects a clear progression in IPv6 adoption and Geofeed distribution compared to the previous data.

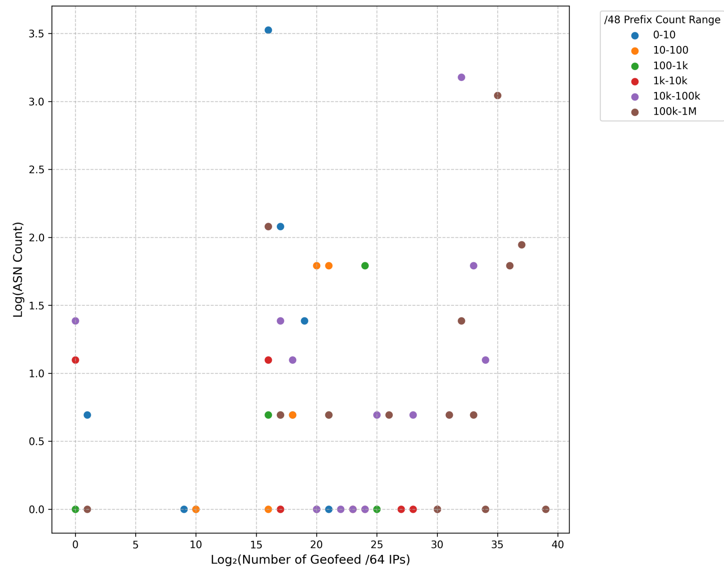

Figure 4: Distribution of ASNs by IPv6 Geofeed coverage percentage from August 2024 snapshot.

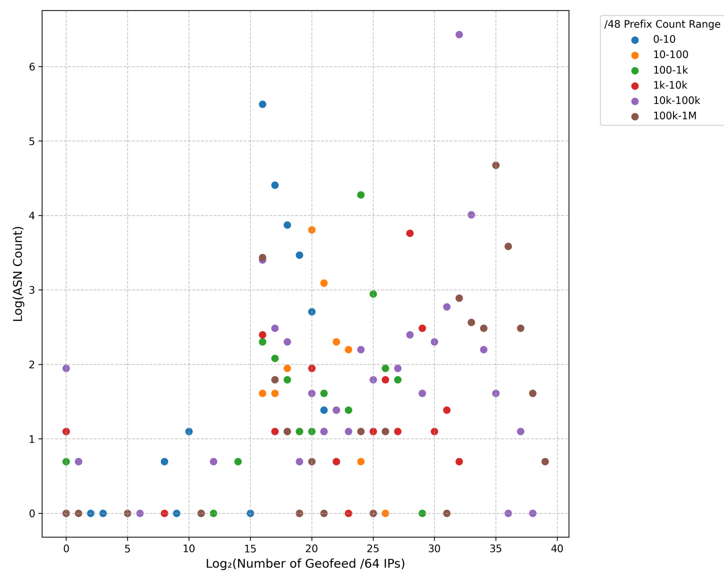

Figure 8: Distribution of ASNs by IPv6 Geofeed coverage percentage from April 7, 2025 snapshot.

6.4.2 Comparison of Invalid Geofeed Data Percentages Within ASNs

Table 15 shows that 92.6% of ASNs providing IPv4 Geofeed data in the August 2024 snapshot submit error-free entries. In the April 7, 2025 snapshot, this figure increases to 94.4%, indicating that ASNs are submitting more valid data and demonstrating improved compliance with RFC standards.

Error Status	Snapshot from August 2024		Snapshot from April 7, 2025	
	ASN Count	Percentage	ASN Count	Percentage
Has No Error	616	92.6%	6736	94.4%
Has at least 1 Error	49	7.4%	401	5.6%

Table 15: Comparison of the count and percentages of ASNs providing at least one IPv4 Geofeed entry.

Figures 5 and 9 illustrate a bimodal error distribution among ASNs providing IPv4 Geofeed data, with peaks at the extreme ends of the error spectrum. The August 2024 snapshot shows high error rates for ranges 0-10% with 38.8% and 90-100% with 40.8%. The April 7, 2025 snapshot shows a general increase in number of ASNs, but the percentages show a broader distribution of error among the middle ranges. This indicates that a greater proportion of ASNs with invalid IPv4 Geofeed data are submitting increasingly larger volumes of invalid entries.

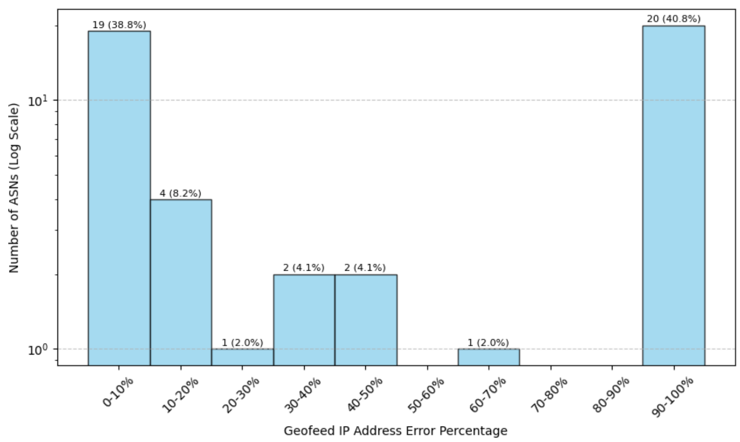

Figure 5: Distribution of errors in the ASNs that has invalid IPv4 Geofeed data from August 2024 snapshot.

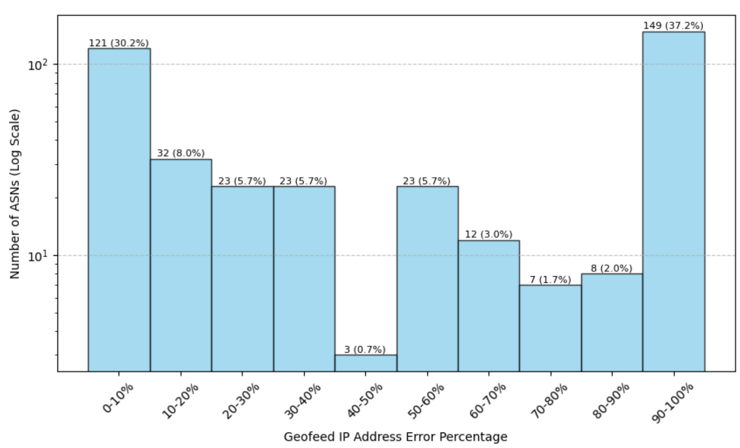

Figure 9: Distribution of errors in the ASNs that has invalid IPv4 Geofeed data from April 7, 2025 snapshot.

Table 16 shows that 85.2% of ASNs providing IPv6 Geofeed data in the August 2024 snapshot submit error-free entries. In the April 7, 2025 snapshot, this figure increases to 92.7%, indicating that ASNs are submitting more valid data and demonstrating improved compliance with RFC standards.

Error Status	Snapshot from August 2024		Snapshot from April 7, 2025	
	ASN Count	Percentage	ASN Count	Percentage
Has No Error	196	85.2%	1803	92.7%
Has at least 1 Error	34	14.8%	141	7.3%

Table 16: Comparison of the count and percentages of ASNs providing at least one IPv6 Geofeed entry.

Figures 6 and 10 show a bimodal error distribution among ASNs providing IPv6 Geofeed data, with prominent peaks at the extreme ends of the error spectrum. While both snapshots show similar proportions with very high error rates (47.1% for August 2024 snapshot and 31.2% for April 7,2025 snapshot in the 90-100% range), the April 7, 2025 snapshot exhibits a broader distribution of errors across the middle ranges. This indicates that a greater proportion of ASNs with invalid IPv6 Geofeed data now exhibit a wider range of data quality as the dataset grows. This rise in invalid data is likely due to the April 7, 2025 snapshot containing approximately 100,000 additional IPv6 Geofeed entries, increasing the likelihood of errors.

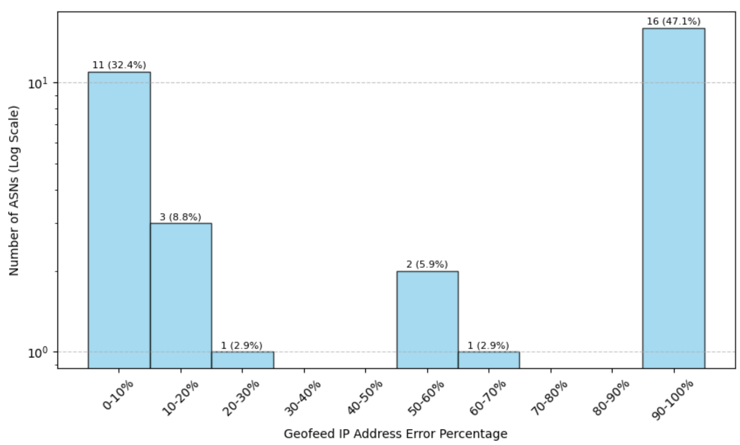

Figure 6: Distribution of errors in the ASNs that has invalid IPv6 Geofeed data from August 2024 snapshot.

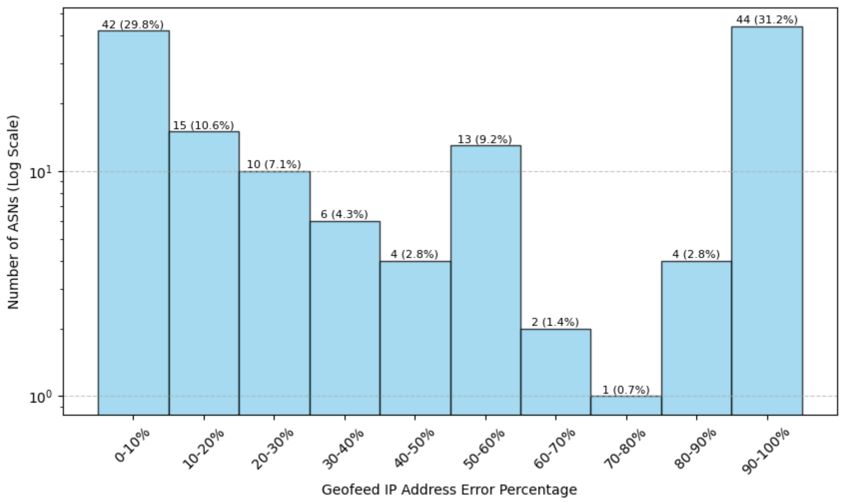

Figure 10: Distribution of errors in the ASNs that has invalid IPv6 Geofeed data from April 7, 2025 snapshot.

7 Future Directions

Our current automation facilitates the validation and analysis of Geofeed data adoption for individual snapshots. However, several important extensions and research opportunities remain.

One significant direction is leveraging automation for longitudinal analysis. Applying the automated system across multiple time-based snapshots would enable the observation of trends in Geofeed adoption and data quality over time. For example, such analysis could reveal whether specific regions or Autonomous System Numbers (ASNs) are improving their compliance with Geofeed standards, and track the evolution of error rates and coverage.

Additionally, the automation can be run using Geofeed data that exists outside of the WHOIS database. While many network operators include Geofeed references in their WHOIS records, some publish Geofeed data independently on the web without referencing it in WHOIS. Relying solely on WHOIS can therefore result in an incomplete view of IP geolocation practices. By incorporating these independently published Geofeeds, the automation enables broader validation and adoption analysis, offering a more comprehensive understanding of how geolocation information is shared across networks.

We also recommend that future validators and tools update their ISO 3166 country and subdivision code databases to include the most current geographic information. Keeping these references up to date will improve the accuracy of validation and reduce false positives due to outdated codes.

Finally, as an extension to our automated process, there is a need for automated mechanisms that can notify corresponding ASes when their Geofeed entries are found to be invalid, including instances of duplicate entries in the WHOIS database. Integrating such notification tools would enable more timely corrections, which could further strengthening the overall integrity of Geofeed information.

8 Comparison with the Previous Findings

In this section, we compare our research findings with the work presented in two previous studies: *You Can Find Me Here: A Study of the Early Adoption of Geofeeds* by Rahel A. Fainchtein and Micah Sherr, and *Geofeeds: Revolutionizing IP Geolocation or Illusionary Promises?* by Ioana Livadariu, Kevin Vermeulen, Maxime Mouchet, and Vasilis Giotsas. Below, we discuss several key differences and advancements made in our research.

8.1 Data Collection and Granularity

Previous Studies: Both prior works relied on biweekly snapshots over limited periods (3.5 months and 14 months, respectively), using the *geofeed-finder* tool to retrieve data from WHOIS records. Livadariu et al. analyzed 73,827 IPv4 and 46,549 IPv6 entries, while Fainchtein et al. focused solely on IPv4 data without specifying entry counts.

Our Study: We collected two snapshots from CAIDA's WHOIS database to analyze the growth of Geofeed entries. The first dataset, retrieved in August 2024, contained 1,169,463 total entries—652,263 IPv4 and 517,200 IPv6. Within this dataset, we identified 84,712 unique Geofeed entries, comprising 60,622 IPv4 and 24,090 IPv6 records. The second snapshot, taken on April 7, 2025, included 1,495,510 entries (902,100 IPv4 and 593,410 IPv6), with 898,204 unique entries—544,766 IPv4 and 353,438 IPv6. This represents a 798% increase in unique IPv4 entries and a 1,368% increase in unique IPv6 entries compared to the August 2024 snapshot, demonstrating significant growth in Geofeed adoption across both protocol versions.

8.2 Validation Rigor and Error Analysis

Previous Studies: Both used *geofeed-finder* for basic RFC compliance checks. Livadariu et al. reported 135 IPv4 and 143 IPv6 entries lacking location data, while Fainchtein et al. noted no errors. Neither classified error types systematically.

Our Study: Before running validation, we removed duplicate Geofeed entries to ensure accurate analysis. We then used geofeed-validator, which categorizes errors into four main types, to assess the quality of the data. Our results showed that over 90% of the errors stemmed from invalid subdivision codes (Table 1). The August 2024 snapshot contained a total of 445 errors, while the April 7, 2025 snapshot showed a significant increase, with 98,049 errors—detailed in Table 7.

8.3 Coverage and Adoption Metrics

Previous Studies: Fainchtein et al. found <1% IPv4 coverage, with industrialized regions overrepresented. Livadariu et al. reported 1.5% IPv4 and 0.7% IPv6 prefix coverage, noting disparities in Africa/Latin America.

Our Study: We quantified adoption at the ASN level using CAIDA's pfx2as tool, revealing:

- IPv4 Geofeed coverage rose from 0.9% to 9.3% of ASNs (Table 9), though coverage rates vary widely across networks. As shown in Figures 3 and 9, ASN coverage exhibits a bimodal distribution—many ASNs either contribute minimal coverage (0–10%) or nearly complete coverage (90–100%). Between August 2024 and April 7, 2025, the proportion of ASNs with minimal coverage dropped from 33.2% to 14.3%, while those with near-complete coverage increased from 23.9% to 38.4%, reflecting a clear shift toward broader adoption and improved Geofeed completeness.
- IPv6 Geofeed coverage rose from 0.7% to 5.6% of ASNs (Table 10), with coverage levels varying across networks. Figures 4 and 10 show a shift from concentrated contributions by large ASNs in August 2024 to a more dispersed distribution in April 2025, indicating broader participation and greater variability in IPv6 Geofeed publishing—though large networks still contribute the bulk of entries.

9 Conclusion

This paper demonstrates that while Geofeeds present a scalable mechanism for enhancing IP geolocation, there remain challenges in adoption and data quality. Our automated validation system shows considerable progress from April 2024 to August 2025: unique IPv4 entries increased by nearly 9 times, from 60,622 to 544,766, and IPv6 entries grew by nearly 15 times, from 24,090 to 353,438. ASN-level interaction also grows, with IPv4 Geofeed coverage expanding by more than 10 times, from 665 to 7,137 ASNs (from 0.9% to 9.3% of the total number of global IPv4 ASNs) and IPv6 coverage growing by more than 8 times, from 230 to 1,944 ASNs (from 0.7% to 5.6% of the total number of global IPv6 ASNs). By comparing these two snapshots, we can better motivate the need for an automated tool to track the evolution of Geofeeds, as it highlights the dynamic changes in Geofeed adoption and helps identify areas for continued improvement.

Despite this increase, data quality issues persist. In the August 2024 snapshot, 445 Geofeed entries fail validation, with 91.2% of these failures due to invalid subdivision codes. By April 2025, validation errors rise to 98,049—entirely due to the same issue, accounting for 96,891 cases (98.8%). These findings underscore the critical need for improved data entry practices and stronger validation mechanisms.

Our approach builds on existing efforts by eliminating duplicates, employing strict RFC-based validation, and tracking daily uptake across the global routing system. The findings show increasing adoption, particularly among large ASNs, but substantial improvements in Geofeed benefits will depend on enhanced standard compliance, broader participation, and continued investment in validation tools and training for network operators.

Acknowledgments

We would like to thank Ricky Mok from CAIDA for his guidance and support in planning our research. He provided us with access to CAIDA's Beamer machine, which included the Geofeed data, and shared the pfx2as tool to help us match IP addresses to the corresponding ASNs.

We also appreciate the guidance of Professor Rocky Chang from Calvin University, whose advice and feedback were instrumental in shaping and directing our research.

References

[1] African Network Information Centre (AFRINIC). AFRINIC: African Network Information Centre. `https://afrinic.net/`, 2025.

[2] American Registry for Internet Numbers (ARIN). About ASN. `https://www.arin.net/resources/guide/asn/`.

[3] American Registry for Internet Numbers (ARIN). ARIN: American Registry for Internet Numbers. `https://www.arin.net/`, 2025.

[4] Asia Pacific Network Information Centre (APNIC). APNIC: Asia Pacific Network Information Centre. `https://www.apnic.net/`, 2025.

[5] BGP Routing Table. Introduction to BGP routing tables. `https://www.bgp.us/routing-table/`.

[6] CAIDA. Center for Applied Internet Data Analysis (CAIDA). `https://www.caida.org/`.

[7] CAIDA Data Server. Index of /datasets/routing/routeviews-prefix2as. `https://publicdata.caida.org/datasets/routing/routeviews-prefix2as/`.

[8] Cloudflight. Geolocation data in marketing strategy. `https://www.cloudflight.io/en/blog/geolocation-data-in-marketing-strategy/`, 2023.

[9] ENSA Project. Active probing. https://ensa.fi/active-probing/#:~:text=Active%20probing%20is%20the%20most,main%20censorship%20infrastructure%2C%20the%20GFW.

[10] Rahel A. Fainchtein and Micah Sherr. You can find me here: A study of the early adoption of geofeeds. *Proceedings of the Passive and Active Measurement Conference (PAM)*, 2024.

[11] ICANN. Internet corporation for assigned names and numbers (ICANN). https://www.icann.org/.

[12] IETF. Internet Engineering Task Force (IETF). https://www.ietf.org/.

[13] Internet Assigned Numbers Authority. Internet assigned numbers authority (IANA). https://www.iana.org/, 2025.

[14] Internet Corporation for Assigned Names and Numbers (ICANN). About WHOIS. https://whois.icann.org/en/about-whois, 2022.

[15] IP Geolocation. IP geolocation API. https://ipgeolocation.io/ip-location-api.html, 2024.

[16] IP Location. Geo-fraud prevention: Using IP geolocation to secure online transactions. https://www.iplocation.net/geo-fraud-prevention-using-ip-geolocation-to-secure-online-transactions, 2023.

[17] IP2Location. ISO 3166-2 subdivision codes. https://www.ip2location.com/free/iso3166-2, 2024.

[18] IP2Location. IP2Location: IP Address to IP Location and Proxy Information. https://www.ip2location.com/, 2025.

[19] ipinfo.io. ipinfo.io: IP address data and geolocation API. https://ipinfo.io/, 2023.

[20] IPXO. What is CIDR? a beginner's guide. https://www.ipxo.com/blog/what-is-cidr/, October 2024.

[21] ISO. ISO 3166 country codes. https://www.iso.org/iso-3166-country-codes.html.

[22] E. Kline et al. A format for self-published IP geolocation feeds, 2020. RFC 8805.

[23] W. Kumari et al. RFC 9092: Finding and using Geofeed data, 2021. RFC 9092.

[24] Latin America and Caribbean Network Information Centre (LACNIC). LACNIC: Latin American and Caribbean Internet Addresses Registry. https://www.lacnic.net/, 2025.

[25] Ioana Livadariu, Kevin Vermeulen, Maxime Mouchet, and Vasilis Giotsas. Geofeeds: Revolutionizing IP geolocation or illusionary promises? *Proceedings of the ACM CoNEXT*, pages 1–21, September 2024.

[26] Massimo Candela. Geofeed-finder. https://github.com/massimocandela/geofeed-finder, 2020.

[27] Massimo Candela. Geofeed-validator. https://github.com/massimocandela/geofeed-validator?tab=readme-ov-file, 2021.

[28] MaxMind. GeoIP databases. https://www.maxmind.com/en/geoip-databases, 2023.

[29] A. McDonald et al. 403 forbidden: A global view of CDN geoblocking. In *Proc. ACM IMC*, 2018.

[30] Number Resource Organization (NRO). About the NRO. https://www.nro.net/about/rirs/.

[31] Philipp Richter et al. A multi-perspective analysis of carrier-grade NAT deployment. In *Proc. ACM IMC*, pages 215–229, 2016.

[32] RIPE NCC. RIR delegation (extended) files. https://ftp.ripe.net/pub/stats/, 2023.

[33] RIPE Network Coordination Centre (RIPE NCC). RIPE NCC: Réseaux IP Européens Network Coordination Centre. https://www.ripe.net/, 2025.

[34] ThousandEyes. What is an autonomous system (AS)? `https://www.thousandeyes.com/learning/glossary/as-autonomous-system`.

[35] ThousandEyes. What is an ISP? (internet service provider). `https://www.thousandeyes.com/learning/glossary/isp-internet-service-provider`.

[36] Brandon Wang et al. Characterizing and modeling patching practices of industrial control systems. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(1):1–23, 2017.

[37] Z. Weinberg et al. How to catch when proxies lie: Verifying the physical locations of network proxies with active geolocation. In *ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2018.

[38] Maya Ziv, Liz Izhikevich, Kimberly Ruth, Katherine Izhikevich, and Zakir Durumeric. ASdb: A system for classifying owners of autonomous systems. *Proceedings of the ACM Internet Measurement Conference (IMC)*, pages 1–17, November 2021.

[39] Zscaler. About EDNS client subnet (ECS) injection. `https://help.zscaler.com/zia/about-edns-client-subnet-ecs-injection#:~:text=Client%20Subnet%20is%20an%20option,be%20passed%20in%20DNS%20requests`.

10 Appendix

GitHub URL: Link to GitHub