You Can Find Me Here: A Study of the Early Adoption of Geofeeds

Rahel A. Fainchtein[1,2](✉) and Micah Sherr[1]

[1] Georgetown University, Washington, DC, USA
raf3272@georgetown.edu
[2] The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA

Abstract. IP geolocation is a popular mechanism for determining the physical locations of Internet-connected devices. However, despite its widespread use, IP geolocation is known to be inaccurate, especially for devices in less industrialized nations. In 2020, *geofeeds* were standardized by the IETF, providing a mechanism for owners of IP addresses (i.e., autonomous systems) to self-report the physical locations of IP blocks under their control. Assuming IP address owners accurately report these locations, geofeeds conceptually have the potential to enable "groundtruth" location data. This short paper takes a first look at the roll-out of geofeeds. We examine the opt-in rates of geofeeds by autonomous systems, and surmise the use of geofeed data by two major IP geolocation providers. Over the course of our 14-month data collection efforts (August 2022–October 2023), the number of IP addresses covered by geofeeds has increased tenfold; however, the adoption rate is still low—less than 1% of the IPv4 address space is covered by geofeeds. We find that the rollout is also uneven, with more industrialized nations opting into geofeeds at rates higher than those of less industrialized ones. Moreover, our comparison of geofeed data to locations reported by commercial IP geolocation services suggests that these commercial services may be beginning to incorporate geofeed data into their resolutions. We discuss the implications of our findings, including the potential that uneven adoption rates may further disenfranchise Internet users in less industrialized nations.

Keywords: IP geolocation · geoblocking · geofeeds

1 Introduction

Determining users' geographic locations has become a source of increased interest to various entities. Inferred location information is regularly used to determine the initial language that should be shown to a user, enable location-based advertising, provide locally relevant news and other geographically-tailored information, enforce usage rights governing copyrighted material, identify potentially fraudulent transactions, restrict access to gambling and other services that have regional restrictions [39] and determine the applicability of laws (e.g., GDPR

P. Richter et al. (Eds.): PAM 2024, LNCS 14538, pp. 228–245, 2024.
https://doi.org/10.1007/978-3-031-56252-5_11

and CCPA), among other uses. Various means of ascertaining or inferring users' locations have become increasingly prevalent across the Internet. In particular, *IP-geolocation*[1], or the estimation of geographic location based on a machine's IP address, has become one of the most common approaches.

Despite the lack of an inherent mapping between an IP address and a geographic location, using IP-geolocation has the distinct advantage of being a technique from which an online user cannot easily opt out. Unlike other sources of users' location information such as HTML headers or mobile phones' GPS coordinates, the IP protocol does not allow a user to make a request without sharing their source address.[2]

In practice, IP-geolocation that supports *geoblocking*, or blocking access to content based on a user's inferred location, and other forms of location-based website customization is generally performed by commercial IP-geolocation services. Though not the only reason, this is due in part to popular websites' and web services' widespread reliance on content distribution networks (CDNs) [9], and CDNs' pervasive use of commercial IP-geolocation services to support ready-made IP-geolocation based offerings, such as CDN-based geofiltering, to their customers [3,6,7,12,21].

Despite the ubiquity of geolocation and geoblocking, commercial geolocation services have been found (1) to be largely unreliable when it comes to geolocating Internet infrastructure (e.g., servers and/or routers) [13,41] and (2) to underrepresent, and more frequently, to mis-locate Internet vantage points in less industrialized nations [37]. Moreover, significant anecdotal evidence suggests that the mis-location or incorrect flagging of IP addresses as VPNs or other geoblock evasion tools is not limited to these applications or vantage points, and that it happens frequently across the Internet (cf., regular and extensive postings in the NANOG listserv complaining about erroneous geolocation of service providers' IP addresses [5]).

Introduced as a means of correcting the issue of mis-location, self-published geolocation feeds or *geofeeds*, were designed to allow network operators such as autonomous systems (ASes) to specify the geographic locations of their IP addresses. The syntax and semantics of geofeeds are codified in an Internet standard [25], which we summarize in the following section. Conceptually, the idea of geofeeds is simple: rather than infer IP geolocation based on heuristics or error-prone triangulation methods, owners of IP addresses announce the geographic locations of their IPs. Further Internet standards govern how these geofeeds can be found and how ASes can update their geofeed records to prevent them from becoming stale [38]. Ideally, geofeeds could serve as "groundtruth" location information, with the important caveat that some IP address owners may have incentives to purposefully report incorrect locations—for example, to support services that attempt to bypass geoblocking [41].

[1] For ease of exposition, we will often use the shorthand *geolocation* to refer to IP-based geolocation.

[2] A user could always use a means of obscuring their IP address such as Tor [16], a VPN, SmartDNS services [18], or a proxy. However, using these technologies requires technical sophistication and imposes performance and usability bottlenecks [17,34].

Geofeeds were standardized in 2020 and began seeing some adoption shortly thereafter. However, to our knowledge, geofeeds have not been studied in the literature. This paper presents the first work towards advancing our understanding of how geofeeds have been adopted. More concretely, this paper attempts to answer the following two research questions:

- **RQ1:** To what extent have geofeeds been adopted by "IP owners"?
- **RQ2:** Is there evidence that suggests that commercial IP-geolocation providers use geofeeds?

Towards answering the above research questions, we performed a measurement study over a 14 month period (August 2022 through October 2023) in which we collected geofeed records and measured their coverage of the IPv4 address space. During this period, we compared geofeed information to two popular IP geolocation services to gauge their level of agreement with geofeed data. We posit that our collection of historical geofeed information may be of independent interest to network researchers, and thus we make our data available at https://github.com/GUSecLab/geofeed-measurement.

Our results show that while geofeed adoption is limited to date (comprising roughly 0.8% of the IPv4 address space), its adoption rate is rapidly increasing—the number of covered IP addresses increased tenfold during our 14-month study. We find that adoption rates are not universal, and that less industrialized countries often (but not always) have smaller adoption rates; this is especially concerning given that countries with less Internet infrastructure experience higher rates of geoblocking by websites based in regions where Internet infrastructure is more readily available [40]. Finally, our results suggest that commercial IP geolocation services may be incorporating geofeed information into their offerings; at the very least, commercial geolocation results tend not to contradict information from geofeeds. Overall, the commercial providers' country-level agreement with geofeeds steadily improved over the course of our measurement period.

2 Background and Related Work

The mechanisms for IP-geolocation can generally be separated into passive and active approaches. Passive IP geolocation techniques rely on querying and parsing data from WHOIS [20] and/or other publicly accessible sources to learn location information. Active approaches, on the other hand, use empirical measurement methods such as multilateration from different Internet vantage points to determine the bounded geographical region in which an IP address must reside [11,19,27,41].

In general, active probing approaches tend to more accurately reflect the state of IP addresses' allocations at time of measurement than passive ones. However, they still pose significant drawbacks including the potential for interference due to high levels of cross traffic and/or network congestion, and regional variation in rates of targets' responsiveness to ICMP and traceroute messages, which are most commonly used for probing [10,36].

In light of these drawbacks, both researchers and industry have turned to more passive geolocation approaches. In particular, for use-cases in which web servers must ascertain incoming connection requests' geographic locations in real time, many rely on commercial IP-geolocation services. Although the exact geolocation methodologies used by commercial IP-geolocation services are often proprietary, previous work has shown that their level of accuracy is inconsistent across different geographic regions and is prone to errors [28,33,37]. Muir et al. and Poesse et al. independently infer that these services, which assemble a database of IP address to geographic location mappings, are likely based off of publicly available resources such as WHOIS, which have been found to have high rates of errors [33,37]. Errors in these databases are often caused by the staleness of source records, sources' lack of data granularity, and their lack of data authentication [33]. In more detail, Dainotti et al. and Richter et al., who evaluate system logs as sources of geolocation information, note that datasets originating in system logs, network flow logs, and/or large-scale network traces must be collected from Internet vantage points that can observe high volumes of traffic, and often contain numerous spoofed addresses, which must then be identified and filtered out [14,36].

Despite the potential obstacles to building accurate IP-geolocation databases, Gharaibeh et al. [28], who measure commercial IP-geolocation services' ability to accurately locate Internet routers, find high rates of country-level interservice agreement across six commercial IP-geolocation services. However, they highlight that these high rates of agreement do not translate into similar rates of accuracy. Specifically, the commercial datasets show between 77.5% - 89.4% country-level accuracy over the routers included within their respective databases—a significantly lower country-level accuracy than these services respectively advertise [28]. Gharaibeh et al. further caution that these services' country-level accuracy varies widely across different countries and regions.[3] For these reasons, Gharaibeh et al. recommend against the use of commercial IP-geolocation for the identification of routers or Internet infrastructure [28].

RFC 8805 [25] introduces self published geolocation information feeds, or *geofeeds*, as a means by which network operators such as ASes or Internet Service Providers (ISPs) can share the actual geolocation data of the IP addresses and prefixes they control. RFC 8805 defines a geofeed file to be a comma separated value (CSV) file in which each entry contains either a single IP address or range of addresses and its corresponding city and country-level geographic information.

RFC 9092 [38] expands on this by centralizing the location to which network operators can publish the URLs of their geofeeds. This enables geofeed consumers, such as commercial IP-geolocation providers, to easily find them. In more detail, RFC 9092 defines a mechanism through which network operators can register their geofeed URL with their respective Regional Internet Registries (RIRs), National Internet Registries (NIRs), or Local Internet Reg-

[3] For example, while most commercial providers showed over 90% accuracy in identifying routers in the U.S., most providers showed between 20% and 39% accuracy when locating routers in Canada [28].

istries (LIRs) and in so doing, add them to their WHOIS database entries. As part of this update, RFC 9092 proposes the expansion of the Routing Policy Specification Language (RPSL) used by Internet Registries to specify registrant information, to include a new geofeed field which holds the URL of the registrant's geofeed [38].

3 Data Collection and Methodology

To answer the research questions from Sect. 1 ("to what extent have geofeeds been adopted by IP owners" and "is there evidence that suggests they are used by commercial IP-location providers?"), we performed a 14-month data collection effort during which we collected all published geofeeds. In what follows, we detail our methodology and data and the limitations of our approach.

3.1 Geofeed Measurement Methodology

To collect geofeed information, we used the open-source geofeed finder [29] tool. The geofeed finder queries current WHOIS records to locate geofeed URLs, pulls the geofeeds' contents, and verifies their integrity in accordance with the requirements set by RFC 8805 [25] and RFC 9092 [38].

Using the tool, we initially queried the geofeed records in April 2022, and, starting in August of that year, pulled updated geofeed records every 13–16 days[4] over the course of a 14 month period (August 2022 - October 2023). We chose to pull (approximately) biweekly to incur only negligible load on WHOIS services.

As of the time of writing, our collection of geofeed data is still ongoing. Our database is effectively a collection of temporal snapshots, each of which reflects the Internet-wide deployment of geofeeds at a moment in time. We anticipate this may be of use to other network and security researchers (e.g., to assess the trustworthiness of geofeeds or to detect equivocation in geofeed records), and make our analysis tools and data available at https://github.com/GUSecLab/geofeed-measurement.

Ethics. We accessed only publicly available information (geofeeds) that are published (publicly) by ASes. The geofeed finder tool accesses bulk WHOIS records for each major RIR and caches the results to minimize the load incurred by repeated querying. As noted above, we further reduce the load incurred by our study by performing the geofeed queries only once every two weeks.

3.2 Limitations

Geofeed-Finder Version Changes. The software version used between April 2022 through January 2023 was deprecated and became unusable during the measurement period and we therefore had to upgrade on February 1, 2023. In June

[4] Measurements were initially pulled manually once every two weeks and were later automated to run on the 13th and 28th of each month.

2023 and September 2023 we once again had to upgrade the geofeed-finder tool since the previous version again became deprecated. While we do not believe that these version changes would significantly affect our results, we were unable to re-pull past geofeeds using the newer software.

Assumption of Geofeed Reliability. In our measurements of geofeeds, we assume the geofeed entries we obtain from the geofeed finder [29] accurately depict the geographic regions to which its publishers (or rather the network operators) allocate the IP addresses under their control.

Additionally, geofeed information is pulled from WHOIS, and is subject to spoofing attacks. As RFC 9092 explains, malicious network operators could exploit the weak or missing authentication of numerous RPSL repositories to spoof `inetnum:` entries and set them to point to geofeed files that contain inaccurate location information [38].

Authenticating geofeeds can be straightforwardly addressed by requiring that all network operators register and publish geofeed files that are digitally signed with their private RPKI keys. Unfortunately, since RPKI is not universally applied, this is unlikely to occur in the near term.

However as a potential stop-gap solution, RFC 8805 and RFC 9092 require geofeed consumers to perform additional checks on geofeed records before consuming them. These checks include verifying that network operators actually control the IP addresses included in their geofeed records, ensuring all consumed geofeeds are transmitted using HTTPS, and that their geofeed parsers process the data in a consistent way [25,38]. In keeping with the standard, we note that the geeofeed finder performs these checks by default and that we intentionally did not configure it to skip them.

4 RQ1: Geofeed Adoption by Network Operators

The expansion of geofeed coverage requires network operators to actively opt-into using geofeeds, and to accurately update them when they reallocate addresses to new geographic regions. Network operators who want their geofeed URLs to be published within their WHOIS (`inetnum:`) entries can register them with their respective Internet Registries. In this section, we investigate how and where networks operators have opted in and the extent to which different geographic locations are represented within the geofeeds.

In measuring geofeed adoption (**RQ1**), we sought to answer the following more specific research questions:

1. At what rate are network operators registering and publishing new geofeeds?
2. What is the coverage of geofeed records across the Internet?
3. How has this coverage changed over time? and
4. Where are the IP ranges listed within these geofeeds geographically located? Are they evenly distributed or concentrated within a few geographic regions?

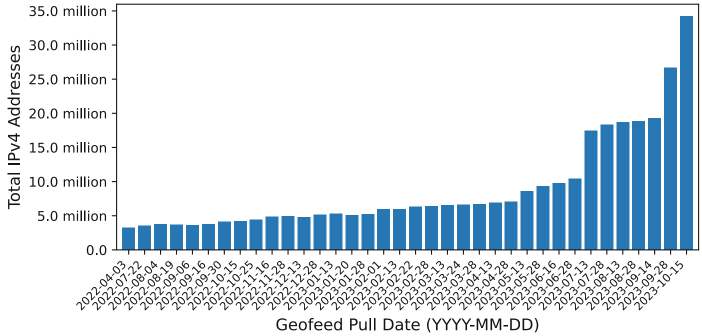

Fig. 1. Number of IPv4 addresses covered by geofeeds.

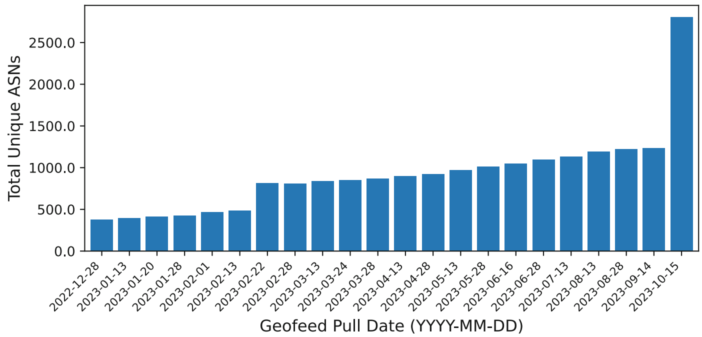

Fig. 2. Unique ASNs that publish geofeeds.

4.1 Geofeed Adoption Has Increased Tenfold in About a Year

Overall, we find network operators' adoption of geofeeds slowly but steadily grew over the course of the observation period. As shown in Fig. 1, geofeeds' IPv4 space coverage grew more than tenfold from 3.22M IPs on April 2, 2022 to 34.3M IPv4 addresses on October 15, 2023. Figure 2 shows similar patterns across the unique ASNs identified in each geofeed pull.

While the number of IPv4 addresses is monotonically increasing (with a few exceptions), there are discernible "bursts" in which large numbers of IPs become covered (see, for example, July 2023 in Fig. 1). This unevenness is largely due to a relatively small number of ASes with large IP blocks opting into publishing geofeeds. Until October 2023, the number of participating ASes has grown more linearly, as can be seen from Fig. 2.

While there has been significant growth in geofeed opt-in, geofeeds' coverage over the Internet address space remains minimal. In particular, as of October 15, 2023, the 34.3M IPv4 addresses announced within the geofeeds only account for roughly 0.80% of the IPv4 address space, or roughly 0.93% of all allocated IPv4 addresses [35]. Similarly, the number of unique ASNs identified increased

Table 1. Geofeed representation by continent.

Continent	Number of IPs	Percent of Gfeed
Europe	13,872,160	40.5%
North America	11,668,093	34.1%
Oceania	7,072,881	20.7%
Asia	725,212	2.61%
South America	214,273	1.91%
Africa	42,664	0.18%
Antarctica	26	0.000076%

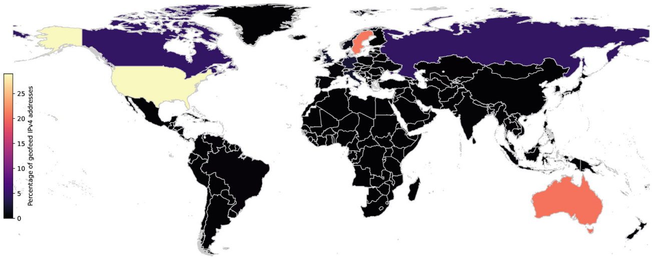

Fig. 3. Percentage of in-country IPv4 addresses covered by geofeeds.

sevenfold from 375 ASNs in December 2022 to 2,805 ASNs in October 2023, but this only equates to 2.4% of the ASNs allocated [35].

4.2 Adoption Rates Vary Significantly by Geographic Region

Despite geofeeds' overall limited opt-in to date, we observe that network operator opt-in for geofeeds varies widely by geographic region. Figure 3, which provides a heatmap of the total IPv4 addresses associated with each country within the geofeeds, shows a strong concentration of the geofeed IPv4 addresses geolocated to wealthier and more industrialized countries that have more Internet infrastructure available.

This is particularly well exemplified by the breakdown of geofeed IPs by location in April 2022 - where IPv4 addresses geolocated to the United States account for about 73.7% of all geofeed IPv4 addresses. While by October 2023 numerous additional countries held more substantial proportions of the total geofeed IPv4 addresses, the vast majority (about 32.6M, or 95.3%) of geofeed IPv4 addresses were located in Europe, Oceania and North America combined. In stark contrast, only 62k IPv4 addresses, or 0.18% of those published in the October 2023, geofeeds were geolocated to Africa. Moreover, among the 20 least represented countries, twelve are located in Africa, and that no countries see an increase in the total geofeed IPv4 addresses geolocated to them.

Assessing the regional breakdown of representation by continent, as given in Table 1, further highlights this trend. Here we find that continents with lower

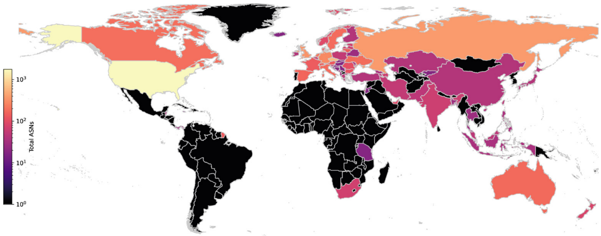

Fig. 4. Total ASNs included in the November 10, 2023 Geofeed Results by country. Note that the coloring is uses a logarithmic scale.

proportions of industrialized countries, or of countries with strong Internet hosting infrastructure, are less represented within the geofeed results.

To account for additional factors we also assess the breakdown of country-level representation in geofeeds normalized by IPv4 address allocation as reported by the NRO [35] and each country's total Internet users as estimated by the CIA World Factbook [8]. Heatmaps of these normalized breakdowns can be found in Fig. 9 and Fig. 10 of Appendix B. Overall we observe a similar correlation between a country's representation within the geofeed results in the normalized results, but note a few key outliers. In particular, we highlight that despite its immense Internet presence, China is the third least represented country in the geofeeds after normalizing by IPv4 address allocation and the twelfth least represented after normalizing by total Internet users. However, while China do not appear to follow the same trend, we suspect additional factors may be contributing to the limited rates at which Chinese IP owners have opted into publishing geofeeds.

AS Opt-in Rates. In addition to looking into the rate of growth over the IPv4 address space, we also sought to gain more insights into the organizations that publish them. In more detail, we wanted to ascertain where the ASes that opted into publishing geoofeeds were geographically located and the types of organizations to which they pertained. Here, we consider a snapshot of geofeeds from November 2023, and use Stanford ASdb [32] from May 2023 to classify ASes.

Figure 4 shows a global heatmap of the geographic distribution of ASes that opted into publishing geofeeds. Similar to Fig. 10, we see that most of the ASes are concentrated in the "Global North" and that countries located in Central and South America, Africa and much of the Middle East are not represented. Additionally, we note that the United States' total opt-in of 1,763 ASes vastly exceeds that of any other country.

When it came to the types of organizations opting into publishing geofeeds, the majority of ASes that opt in appear to be Internet Service Providers (ISPs). Table 2a shows the breakdown of the top ten most represented categories of ASes that participate in geofeeds as determed by ASdb [32]; we further detail the subcategories for "Computer and Information Technology" in Table 2b. As shown in Table 2a, roughly 1,927 ASes, or 72.7% of the 2,652 covered by the ASdb,

Table 2. Categorical Breakdowns of ASes

Category	N.ASNs	%ASNs
Computer & Info. Tech.	1927	72.7%
Other	451	17.0%
Service	446	16.8%
Retail,Wholesale & E-comm.	251	9.5%
Finance & Insurance	237	8.9%
Media,Publishing & Broadcast	226	8.5%
Construction & Real Estate	170	6.4%
Education & Research	101	3.8%
Manufacturing	91	3.4%
Government & Public Admin.	80	3.0%

(a) Categories of ASes that publish geofeeds.

Subcategory	No. ASNs
Internet Service Provider (ISP)	1634
Hosting and Cloud Provider	389
Software Development	380
Computer and Network Security	68
Technology Consulting Services	64
Phone Provider	56
Other	20
Search	15
Satellite Communication	2

(b) Breakdown of ASes in *Computer and Information Technology* category.

are identified as Computer and Information Technology organizations. Moreover, 1,634 or 84.3% of these ASes are denoted as ISPs—accounting for 61.6% of the ASdb covered ASes overall. Given the motivation for introducing/standardizing geofeeds noted in RFC 8805 [25], this is not entirely surprising.

In an effort to gain additional insight into ASes' motivations for opting-into publishing geofeeds, we also examine the geographic distribution of ISPs across the ASNs that publish geofeeds. Figure 11 of Appendix B shows a heat-map of the geographic distribution of ISPs. Note that coloring here is in logarithmic scale to show the full variance. Here once again we see an absence of most countries in South America, Africa and the Middle East and the largest concentration of ISPs in the United States. Figure 12 of Appendix B shows a heat-map of the proportion of each countries' ASes that are ISPs. Among the countries from which ASes opted to publish geofeeds, a larger proportion of the ASes in countries with fewer geofeed ASes were ISPs.

Given geofeeds' intended role of preempting IP ranges' geographic mislocation by commercial IP-geolocation services, this discrepancy could further contribute to existing regional disparities in users' Internet accessibility and QoS if it continues to perpetuate. We discuss this in more detail in Sect. 6.

5 RQ2: Evidence of Commercial Adoption of Geofeeds

To study geofeeds' impact on the accuracy of commercial IP-geolocation services (**RQ2**), we compare geolocation estimates from two of the most popular commercial IP geolocation services [4], Maxmind-GeoIP2 and IPgeolocation.io, to the geolocation information given in the geofeeds. Since it is supplied directly from the owners of its covered IP addresses, we consider the geofeed data as "ground truth" and evaluate the degree to which the commercial IP-geolocation services agree with it. We show the geofeed fetch dates and those of corresponding commercial database (DB) accesses for Maxmind and IPgeolocation.io in Table 3 in Appendix A.

We performed the following steps to assess the agreement between the geofeeds and the commercial IP geolocation datasets:

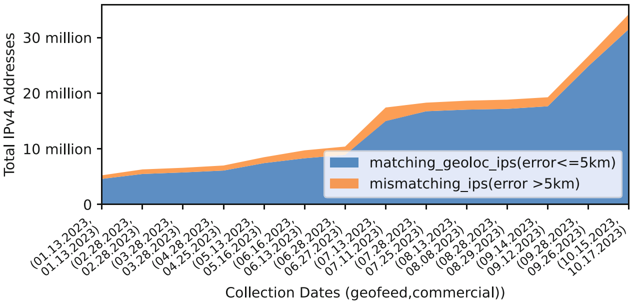

Fig. 5. Maxmind overlap with geofeed and overall accuracy.

1. We identified pairings of geofeed and commercial DB records where the IPv4 addresses referenced in their respective CIDR prefixes overlapped.
2. We compared the location named in each pairing's geofeed record with the one named in the commercial DB's estimate.[5]
3. In cases where the location names did not match, we then performed reverse-geocoding on the geofeed location names to ascertain their approximate geographic coordinates. Since Maxmind and IPgeolocation.io both reported using the Geonames reverse geocoding database [1] for this task [22,30], we decided to use it as well to maintain consistency.[6]
4. Using the geofeed locations' estimated geographic coordinates, we then computed the approximate geodesic distance [2,24] between them and the (corresponding) commercial entries' location estimates.

Upon completing these steps, we then assessed the overall agreement on both the country and city levels across all IPv4 addresses that were included in both the geofeed result and the contemporary commercial DB pull.

Results. The agreement between (1) geofeeds and (2) Maxmind-GeoIP2 and IPgeolocation.io could suggest that these services may consult geofeeds when updating their (respective) IP-geolocation DBs. As shown in Figs. 5 and 7, there are very high rates at which both commercial providers provided geolocation estimates for the corresponding geofeed's IPv4 addresses and their high accuracy rates on the geolocation of these addresses. Both commercial providers covered 99% – 99.9% of their contemporary geofeeds' IPv4 addresses. Moreover, their country and city-level accuracy for geofeed IPs roughly met or significantly surpassed their respective self-reported accuracy rates.

[5] To account for locales having numerous names or versions of the same name (e.g., the city name for Đakovo, Croatia could also be spelled Djakovo or Dakovo), we computed the normalized Damerau-Levenshtein distance [15] between the two location names and asserted that to match, the result had to be less than 0.5.

[6] To account for locations having multiple names or spellings, we used fuzzy matching with tokenized Levenshtein distance to find many of the named locations.

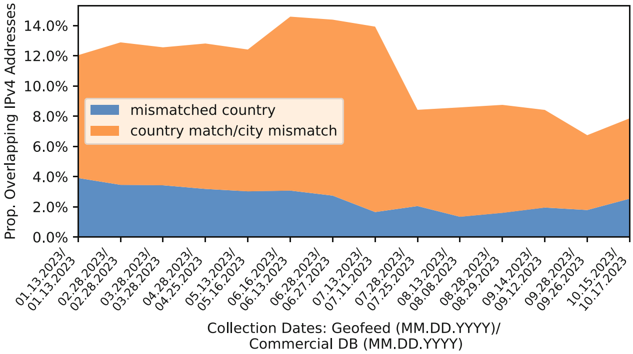

Fig. 6. Maxmind inaccuracy rates compared to geofeed results.

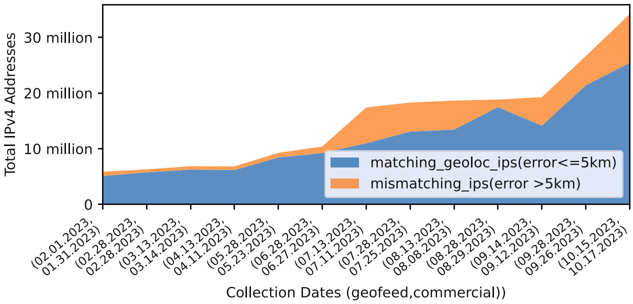

Fig. 7. IPgeolocation.io's overlap with geofeeds and overall accuracy.

At the country level, the observed agreement between geofeed results and commercial estimates was about 96% and 98% for Maxmind-GeoIP2 and IPgeolocation.io respectively, and very close to their self-reported accuracy rates of 99% (Maxmind) and 99.9% (IPgeolocation.io) [23,31]. As shown in Figs. 6 and 8, city-level agreement between geofeed results and commercial estimates ranged from 85.1%–93.1% (mean = 88.5%) for Maxmind and between 63.0%–92.7% (mean = 81.7%) for IPgeolocation.io. It is worth noting that both providers' average rates of city level geofeed agreement exceed their self-reported accuracy rates of 66% within a 50km radius for Maxmind-GeoIP2 and 75% for IPgeolocation.io [23,31].

6 Discussion

Our findings indicate a potentially increasing correlation between a country's level of industrialization and its overall representation within the geofeed results. If network operator opt-in to geofeeds becomes the norm in more industrialized locations, our observations indicate that the accuracy with which commercial IP-geolocation providers would be able to locate a given IPv4 Internet vantage point could become increasingly correlated with the extent to which the

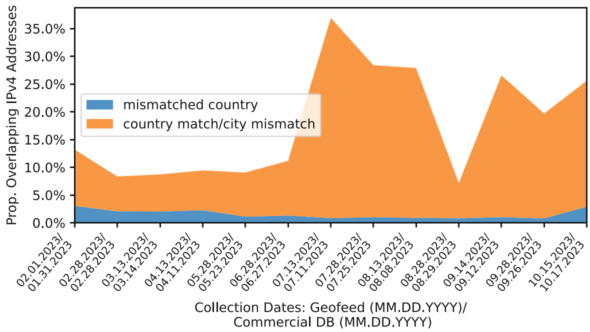

Fig. 8. IPgeolocation.io inaccuracy rates compared to geofeed results.

country housing it is industrialized. This is particularly concerning in light of existing works showing that Internet connections originating in less industrialized countries face higher rates of geoblocking [40] and that if given the option, websites/host servers will likely geofilter traffic in cases where doing so is not actually necessary [9,26]. Since websites and online services frequently rely on commercial IP-geolocation service responses to dictate who can access them and how they behave once accessed, this would translate into a growing discrepancy in the accuracy with which web hosts would be able geolocate Internet vantage points based on their rates of industrialization. Moreover, it implies that less industrialized countries could sustain further degradation to their overall Internet access and QoS as a result.

We are continuing to collect geofeed records, and are in the process of preparing an open (free) repository to house historical geofeed information[7]. We believe that this will be a valuable resource both for researchers interested in geofeeds in particular, and more generally, for those wishing to understand how network operators opt-in to new Internet standards.

Acknowledgments. We thank the anonymous reviewers and shepherd for their invaluable feedback and suggestions. This work is partially funded by the National Science Foundation through grants 1925497 and 2138078, and by the Callahan Family Chair fund. The opinions and findings expressed in this paper are those of the authors and do not necessarily those of any employer or funding agency.

[7] See https://github.com/GUSecLab/geofeed-measurement.

A Geofeed and Commercial IP Fetch Dates

Table 3 lists the dates of fetches for the geofeeds and the corresponding dates of the commercial IP datasets that were used for comparison.

Table 3. Mapping of pull dates for geofeed results and matched commercial DB pulls. Pairings were selected to minimize the time between the geofeed and commercial pull dates (or vice versa).

Gfeed Date	Maxmind Date	Gfeed Date	IPgeoloc. Date
2023-01-13	2023-01-13		
2023-02-28	2023-02-28	2023-02-01	2023-01-31
2023-03-28	2023-03-28	2023-02-28	2023-02-28
2023-04-28	2023-04-25	2023-03-13	2023-03-14
2023-05-13	2023-05-16	2023-04-13	2023-04-11
2023-06-16	2023-06-13	2023-05-28	2023-05-23
2023-06-28	2023-06-27	2023-06-28	2023-06-27
2023-07-13	2023-07-11	2023-07-13	2023-07-11
2023-07-28	2023-07-25	2023-08-13	2023-08-08
2023-08-13	2023-08-08	2023-08-28	2023-08-29
2023-08-28	2023-08-29	2023-09-14	2023-09-12
2023-09-14	2023-09-12	2023-09-28	2023-09-26
2023-09-28	2023-09-26	2023-10-15	2023-10-17
2023-10-15	2023-10-17		

B Country's Representation Within the Geofeed Results

Table 4 presents a breakdown of the top ten most and bottom 20 least represented countries within the geofeed results before normalization.

Figure 9 provides a breakdown of countries' representation within the geofeeds normalized by their respective number of Internet users [8] and Fig. 10 shows geofeeds normalized by each country's IPv4 address allocation. Additionally, Fig. 11 provides a country-wise breakdown of the total ASes categorized as ISPs by the ASdb in the November 10, 2023 geofeed results and Fig. 12 denotes the proportion of ISPs amongst each represented country's ASes in the same geofeed data.

Table 4. Top ten most *(top)* and bottom 20 least *(bottom)* represented countries.

Country	Continent	IPs (Apr22)	% Gfeed (Apr22)	IPs (Oct23)	% Gfeed (Oct23)
United States	North America	2,374,878	73.75%	9,919,174	29.0%
Sweden	Europe	4,683	0.145%	7,106,676	20.8%
Australia	Oceania	5,208	0.162	6,992,481	20.43%
Canada	North America	158,476	4.92%	1,597,552	4.67%
Russia	Europe	16,000	0.497%	1,568,923	4.58%
Germany	Europe	29,815	0.926%	844,017	2.47%
United Kingdom	Europe	88,158	2.74%	691,188	2.02%
Denmark	Europe	184,138	5.72%	632,484	1.85%
Italy	Europe	15,549	0.483%	630,185	1.84%
Netherlands	Europe	87,242	2.71%	477,869	1.40%
Djibouti	Africa	10	0.000 311%	6	0.000 023%
Chad	Africa	10	0.000 311%	6	0.000 018%
Mauritania	Africa	10	0.000 311%	6	0.000 018%
Nauru	Oceania	8	0.000 248%	6	0.000 018%
Sudan	Africa	8	0.000 248%	6	0.000 018%
Senegal	Africa	8	0.000 248%	6	0.000 018%
Sao Tome and Principe	Africa	8	0.000 248%	6	0.000 018%
Holy See (Vatican City)	Europe	8	0.000 248%	6	0.000 018%
Marshall Islands	Oceania	6	0.000 186%	6	0.000 018%
Niger	Africa	6	0.000 186%	6	0.000 018%
Tonga	Oceania	6	0.000 186%	6	0.000 018%
Samoa	Oceania	6	0.000 186%	6	0.000018%
Dominica	North America	8	0.000 248%	4	0.000012%
Mali	Africa	8	0.000 248%	4	0.000012%
Nicaragua	North America	8	0.000 248%	4	0.000012%
Burundi	Africa	6	0.000 186%	4	0.000012%
Comoros	Africa	6	0.000 186%	4	0.000012%
Togo	Africa	6	0.000 186%	4	0.000012%
Uganda	Africa	6	0.000 186%	4	0.000012%
Cuba	North America	6	0.000 186%	2	0.000 006%

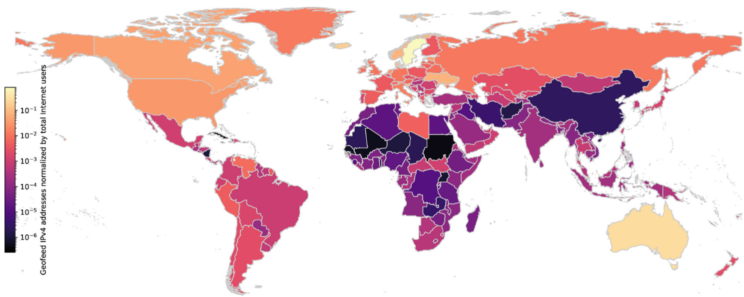

Fig. 9. Countries' IPv4 address representation within the geofeeds normalized by number of Internet users [8].

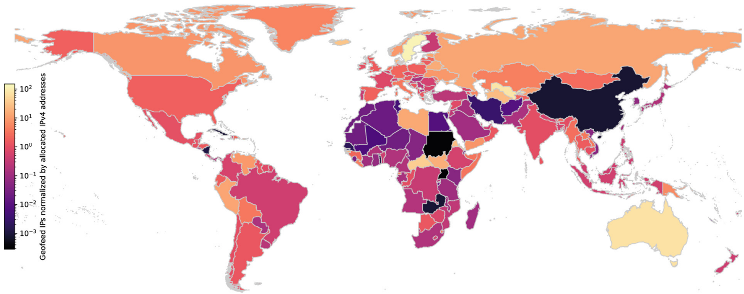

Fig. 10. Countries' IPv4 address representation within the geofeeds normalized by their IPv4 address allocations.

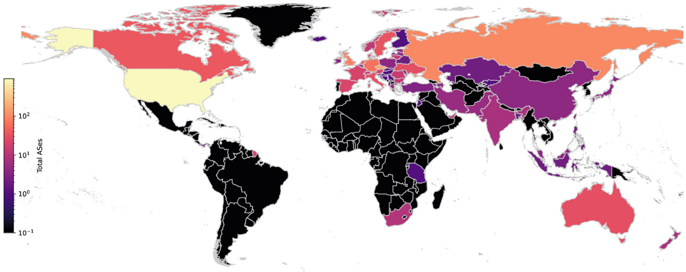

Fig. 11. Total number of ISPs in each country within the Geofeed Results for November 10, 2023 as categorized by the Stanford ASdb.

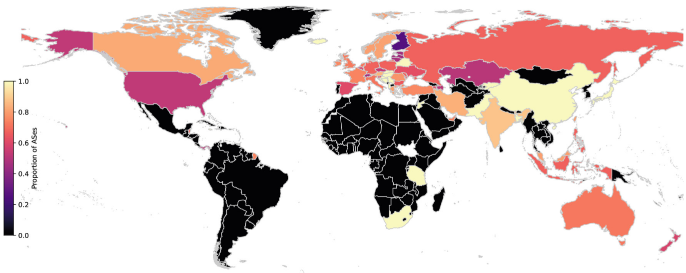

Fig. 12. Proportion of each country's ASNs that were categorized as ISPs by the Stanford ASdb in the November 10, 2023 Geofeed Results.

References

1. Geonames. https://download.geonames.org/export/dump/
2. Geopy documentation. https://geopy.readthedocs.io/en/latest/
3. IP geolocation database: Fastly Support. https://support.fastly.com/hc/en-us/community/posts/360078589732-IP-geolocation-database
4. CDN Comparison - Most Popular CDNs of 2021 (2021). https://www.experte.com/website/cdn

5. Seclists.org - re: Google IP geolocation (2021). https://seclists.org/nanog/2021/Apr/67
6. Digital Element (2022). https://www.digitalelement.com
7. Geoguard (2022). https://www.geocomply.com/products/geoguard/
8. Central Intelligence Agency: CIA world factbook. https://www.cia.gov/the-world-factbook/field/internet-users/country-comparison/
9. McDonald, A., et al.: 403 forbidden: a global view of CDN geoblocking. In: Proceedings of the Internet Measurement Conference 2018. IMC 2018. Association for Computing Machinery (2018)
10. Bano, S., et al.: Scanning the internet for liveness. SIGCOM Comput. Commun. Rev. **48**(2), 2–9 (2018)
11. Wong, B., Stoyanov, I., Sirer, E.G.: Octant: a comprejensive framework for the geolocalization of internet hosts. In: 4th USENIX Symposium on Networked Systems Design & Implementation, NSDI 2007. USENIX Association (2007)
12. Cloudflare: Configuring Cloudflare IP Geolocation (2021). https://support.cloudflare.com/hc/en-us/articles/200168236-What-does-Cloudflare-IP-Geolocation-do-. Accessed 6 Dec 2021
13. Iordanou, C., Smaragdakis, G., Poese, I., Laoutaris, N.: Tracing cross border web tracking. In: Proceedings of the Internet Measurement Conference 2018, IMC 2018. Association for Computing Machinery (2018)
14. Dainotti, A., et al.: Estimating internet address space usage through passive measurements. SIGCOMM Comput. Commun. Rev. **44**(1), 42–49 (2014)
15. Damerau, F.J.: A technique for computer detection and correction of spelling errors. Commun. ACM **7**(3), 171–176 (1964). https://dl.acm.org/doi/10.1145/363958.363994
16. Dingledine, R., Mathewson, N., Syverson, P.: Tor: the second-generation onion router. In: USENIX Security Symposium (USENIX) (2004)
17. Fainchtein, R.A., Aviv, A.J., Sherr, M.: User perceptions of the privacy and usability of smart DNS. In: Proceedings of the 38th Annual Computer Security Applications Conference, ACSAC 2022, pp. 591–604. Association for Computing Machinery (2022). https://dl.acm.org/doi/10.1145/3564625.3567978
18. Fainchtein, R.A., Aviv, A.J., Sherr, M., Ribaudo, S., Khullar, A.: Holes in the geofence: privacy vulnerabilities in "smart" DNS services. In: Proceedings on Privacy Enhancing Technologies (PoPETS) (2021)
19. Gueye, B., Ziviani, A., Crovella, M., Fdida, S.: Constraint based geolocation of internet hosts. In: Proceedings of the 2004 Internet Measurement Conference, IMC 2004. Association for Computing Machinery (2004)
20. Internet Corporation for Assigned Names and Numbers: About WHOIS (2022). https://whois.icann.org/en/about-whois
21. IPGeolocation: ip-geolocation-api-jquery-sdk (2022). https://www.jsdelivr.com/package/npm/ip-geolocation-api-jquery-sdk
22. ipgeolocation.io: IP geolocation API and IP lookup documentation. https://ipgeolocation.io/documentation.html
23. ipgeolocation.io: IP Geolocation FAQs. https://ipgeolocation.io/faq.html
24. Karney, C.F.: Algorithms for geodesics. J. Geodesy **87**, 43–55 (2013)
25. Kline, E., Duleba, K., Szamonek, Z., Moser, S., Kumari, W.: A format for self-published IP geolocation feeds. Informational 8805, RFC-Editor (2020)
26. Kumar, R., Virkud, A., Raman, R.S., Prakash, A., Ensafi, R.: A large-scale investigation into geodifferences in mobile apps. In: 31st USENIX Security Symposium (USENIX Security 2022). USENIX Association (2022). https://www.usenix.org/conference/usenixsecurity22/presentation/kumar

27. Laki, S., Mátray, P., Hága, P., Sebők, T., Csabai, I., Vattay, G.: Spotter: a model based active geolocation service. In: 2011 Proceedings IEEE INFOCOM (2011)
28. Gharaibeh, M., Shah, A., Huffaker, B., Zhang, H., Ensafi, R., Papadopoulos, C.: A look at router geolocation in public and commercial databases. In: Proceedings of the Internet Measurement Conference 2017, IMC 2017. Association for Computing Machinery (2017)
29. Massimo Candela: geofeed-finder (2022). https://github.com/massimocandela/geofeed-finder
30. Maxmind: Geoip2 and geolite city and country databases. https://dev.maxmind.com/geoip/docs/databases/city-and-country
31. Maxmind: Geolocation accuracy. https://support.maxmind.com/hc/en-us/articles/4407630607131-Geolocation-Accuracy
32. Ziv, M., Izhikevich, L., Ruth, K., Izhikevich, K., Durumeric, Z.: ASdb: a system for classifying owners of autonomous systems. In: ACM Internet Measurement Conference (IMC) (2021)
33. Muir, J.A., Oorschot, P.C.V.: Internet geolocation: evasion and counterevasion. ACM Comput. Surv. **42**(1), 1–23 (2009)
34. Namara, M., Wilkinson, D., Caine, K., Knijnenburg, B.P.: Emotional and practical considerations towards the adoption and abandonment of VPNs as a privacy-enhancing technology. Proc. Priv. Enh. Technol. (PoPETS) **2020**(1), 83–102 (2020)
35. NRO: Delegated-extended-file. https://ftp.ripe.net/pub/stats/ripencc/nro-stats/latest/nro-delegated-stats. Accessed 15 Oct 2023
36. Richter, P., Smaragdakis, G., Plonka, D., Berger, A.: Beyond counting: new perspectives on the active IPv4 address space. In: Proceedings of the 2016 Internet Measurement Conference, IMC 2016. Association for Computing Machinery (2016)
37. Poese, I., Uhlig, S., Kaafar, M.A., Donnet, B., Gueye, B.: IP geolocation databases: unreliable? ACM SIGCOMM Comput. Commun. Rev. **41**(2), 53–56 (2011)
38. Kumari, W., Bush, R., Candela, M., Housley, R.: RFC 9092 Finding and Using Geofeed Data. RFC Proposed Standard 9092, Internet Engineering Task Force (IETF) (2021)
39. Trimble, M.: The future of cybertravel: legal implications of the evasion of geolocation. Fordham Intell. Prop. Media Entertain. Law J. **22**, 567 (2012)
40. Tschantz, M.C., Afroz, S., Sajid, S., Qazi, S.A., Javed, M., Paxson, V.: A bestiary of blocking: the motivations and modes behind website unavailability. In: 8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 2018). USENIX Association (2018). https://www.usenix.org/conference/foci18/presentation/tschantz
41. Weinberg, Z., Cho, S., Christin, N., Sekar, V., Gill, P.: 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)