Geofeed Adoption and Authentication

Dipsy Desai, Kicho Yu, Sulyab Thottungal Valapu
University of Southern California, Los Angeles, CA
{deepakde, kichoyu, thottung}@usc.edu

Abstract—IP Geofeed is a recently proposed informational standard that allows network operators to publish the geographical location of deployed IPv4 and IPv6 prefixes. In this work we study the adoption of IP geofeed, assess deployment of geofeed at Regional Internet Registry and Autonomous System levels, and analyze adherence to RFC 8805 and RFC 9092 in deployed geofeeds. We evaluate the authentication mechanism proposed in RFC 9092 and find that it lacks key features from a security perspective. We propose a novel approach to simplify the authentication of geofeeds and assess its efficiency using different benchmarks. Our findings highlight the challenges in current geofeed adoption and the potential for improving both security and scalability in geofeed validation processes.

Index Terms—IP Geofeed, Regional Internet Registry, Autonomous Systems, Authentication of Geofeed

I. INTRODUCTION

IP geolocation is widely recognized for its ability to enhance user experience. For example, it enables selecting geographically closer servers to reduce latency, tailoring user experience based on location (like language or currency), and delivering context-aware search results (such as "events this weekend"). Current state-of-the-art solutions employ a variety of techniques, including multilateration based on latency measurements [1] to estimate the approximate geolocation of hosts, and more recent advancements like machine learning models [2] or search engine clicks [3] to improve IP geolocation.

Despite these advancements, IP geolocation alone is not sufficient to track real-time changes in IP address assignments, highlighting the need for a mechanism to proactively signal such changes. For example, these methods can become temporarily *stale* due to changes in prefix deployments, like Internet Service Provider (ISP) renumbering. To address this, the IETF introduced *geolocation feeds*, or *geofeed*, a technique enabling network operators to publish the geolocation of deployed IP prefixes. The standard is split across two RFCs: RFC 8805 [4] defines the format of geofeed files, while RFC 9092 [5] details the standard methods for publishing and utilizing geofeed files.

Given its status as a nascent technology in current Internet standards, there is limited information on the adoption of geofeed by network operators. In this work, we perform an initial measurement study of geofeed adoption at multiple levels of Internet organization, specifically the Regional Internet Registry (RIR) and Autonomous System (ASes). We analyze the extent to which geofeeds adhere to RFC standards and explore the shortcomings of the *authentication* procedures described in RFC 9092. Based on our analysis, we propose a novel, secure, and scalable two-step authentication method for geofeed publication and validation.

II. ANALYSING GEOFEED ADOPTION

RFC 9092 [5] specifies that geofeed information should be included in the `inetnum`, `inet6num`, or `NetRange` database classes as defined by the Routing Policy Specification Language (RPSL) [6]. To gather geofeed data, our first step was to collect the relevant `inetnum`, `inet6num`, or `NetRange` records from all five RIRs, namely AFRINIC, APNIC, ARIN, LACNIC and RIPE NCC, during February and March 2024. These records were queried from publicly available RIR databases.

Once the records were obtained, we parsed them to extract the URLs of the `csv` files containing the actual geofeed data. These geofeed URLs were then used to download the CSV files associated with each RIR's data. During this phase, we encountered some challenges, for example, approximately 7.76% of the geofeed URLs were inaccessible due to DNS resolution failures, connection timeouts, or various HTTP errors. Specifically, out of the total 1547 URLs queries, 1427 were accessible, while the remaining URLs failed to establish connections (the most common issue), or resulted in HTTP 404 (Not Found) errors.

After successfully gathering the data, we proceeded with a detailed analysis aimed at answering several key research questions related to geofeed adoption, compliance with RFC standards, and the efficacy of geofeed authentication methods.

A. Do certain RIRs adopt geofeed more quickly than others?

To better understand the deployment of geofeed, we first look at its adoption at the RIR level. Table I shows the per-RIR breakdown of `inetnums` and `inet6nums` having geofeed information. Noticeably, RIPE NCC leads significantly in the count of both `inetnums` and `inet6nums` with geofeed entries among other RIRs, as well as the number of ASes with geofeed entries. In fact, RIPE accounts for 82.04% and 88.24% of geofeed-enabled `inetnums` and `inet6nums` respectively. However, overall, only 0.25% of `inetnums` and 0.30% of `inet6nums` have associated geofeed entries. This indicates that geofeed adoption is still in its early stages.

B. Do specific categories of ASes adopt geofeed more rapidly?

Since the goal of geofeed is to aid in geolocating IP addresses, different categories of autonomous systems may have different levels of interest in its adoption. For example,

RIR	inetnum		inet6num		# AS
	Count	Fraction	Count	Fraction	
AFRINIC	421	0.28%	24	0.07%	19
APNIC	871	0.07%	141	0.14%	156
ARIN	1375	1.84%	206	0.29%	440
LACNIC	58	0.01%	16	0.06%	21
RIPE	12447	0.30%	2905	0.34%	1417
Total	**15172**	**0.25%**	**3292**	**0.30%**	**1907**

TABLE I

NUMBER AND FRACTION OF INET[6]NUMS PER RIR WITH GEOFEED RECORDS, AND NUMBER OF ASES WITH AT LEAST ONE GEOFEED RECORD. NOTE THAT THE SAME AS MAY GET COUNTED UNDER MULTIPLE RIRS BASED ON AVAILABLE RECORDS.

Fig. 1. Category-wise breakdown of ASes that have geofeed records, grouped by RIR.

ISPs may be motivated to use geofeeds to signal changes in the geographical deployment of customer prefixes, whereas educational institutions may be less inclined to do so since their locations are usually fixed. Hence, it is worth examining whether specific categories of ASes have been adopting geofeed more rapidly than others.

To answer this question, we used the AS information API provided by ipinfo.io to sort geofeed-enabled ASes within each RIR into ISP, Business, Hosting, and Education categories. The results, presented in Figure 1 above, indicate that ASes belonging to the Business category lead geofeed adoption, followed by ISP and Hosting.

Our analysis primarily focuses on the RIR and AS level due to the data's higher-level structure. The records we gathered from the RIR databases include information about IP address ranges assigned to entire ASes or RIRs, but do not provide the level of detail found at the individual IP prefix level. As such, our analysis was constrained by the granularity of the available data. While this high-level analysis offers valuable insights into geofeed adoption and compliance, we recognize that examining geofeeds at a more granular level, such as individual IP prefixes (e.g., /32 for IPv4 or /64 for IPv6), could yield a more precise understanding of geofeed accuracy and adoption. Future work could explore this, if data at this finer granularity becomes available. For a visual representation of our findings, see Appendix B, which highlights the global and some regional geofeed adoption patterns.

C. Do geofeeds adhere to RFCs?

RFCs 8805 and 9092 specify strict requirements for formatting and publishing geofeed. These requirements are necessary to ensure the integrity of geofeeds as well as provide structural uniformity for parsers. For our analysis, we selected a subset of the requirements specified by the RFCs that we identified to be "important" from the perspective of a consumer of geofeed data. We now discuss the results of our RFC adherence analysis.

1) RFC 9092: RFC 9092 is concerned with publishing and discovering URLs of geofeed csv files. We consider adherence of published geofeeds to the following specifications:

1) Geofeeds in inet[6]num entry using the remarks: attribute must be formatted as follows (Geofeed is case sensitive): remarks: Geofeed https://example.com/geofeed.csv
2) Geofeeds in inet[6]num entry using the geofeed: attribute must be formatted as follows: geofeed: https://example.com/geofeed.csv
3) Geofeed URLs must use https.

The RFC also mentions that apart from using remarks:, geofeeds can be also published in a geofeed: attribute of the inetnum or NetRange object once RPSL supports that attribute, however, we have not found any in support of this.

If an inet[6]num passes all three checks, we categorize it as "valid". Otherwise, we categorize it as "invalid formatting" or "not HTTPS" based on which check(s) it failed. The analysis results are illustrated in Fig. 2. In particular, we note that 100 (0.54%) inet[6]nums publish geofeed csvs over the unsecure http protocol.
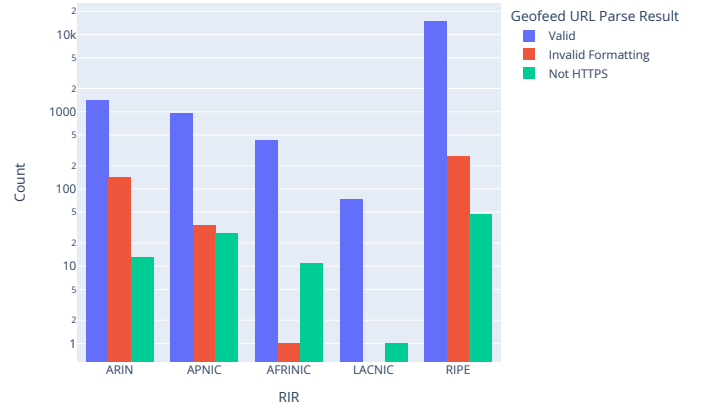

Fig. 2. Results of RFC 9092 adherence analysis.

2) RFC 8805: RFC 8805 is concerned with the content of the geofeed csv files. We consider the following specifications for our analysis:

1) Geofeed csv files must use UTF-8 character encoding and CRLF line breaks.
2) All lines in a geofeed csv file must contain the following five fields in a comma-separated (no-spaces) format: ip_prefix,alpha2code,region,city, postal_code

3) All fields except `ip_prefix` can be empty, but the requisite number of commas must be present.
4) The `ip_prefix` field must be either a single IP address or an IP prefix in CIDR notation. The `alpha2code` and `region` fields, if non-empty, must be ISO country or region codes conforming to ISO 3166-1 alpha 2 and ISO 3166-2 respectively.

If a `csv` line fails any of the check(s), we categorize it as "malformed". We find that 511035 lines (89.51%) out of 570909 are valid. It is particularly alarming that 10.49% of all geofeed lines are malformed and hence unusable. For a closer look at the reasons for malformed lines, we further categorize malformed lines into "not enough fields", "malformed IP prefix", "malformed country code" and "malformed region code" based on which check(s) it failed. The results, shown in Fig. 3, indicate that not having enough fields, malformed IP prefix and malformed region codes are the major reasons for malformed lines. We also find that, although all geofeed files use UTF-8 encoding, only 393 out of 1427 (27.54%) use CRLF line breaks.
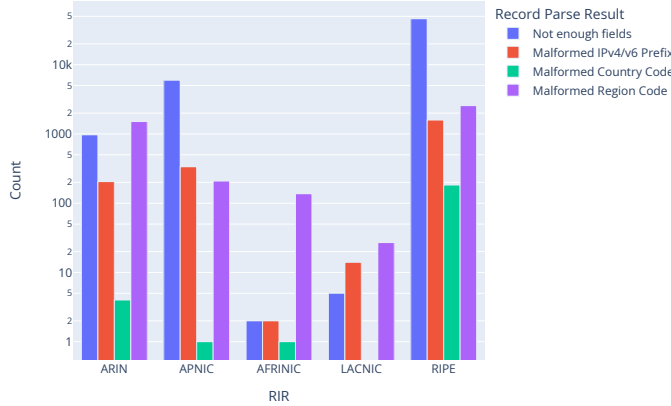

Fig. 3. Results of RFC 8805 adherence analysis.

III. AUTHENTICATING GEOFEEDS

Geofeed data, which provides critical information about the geolocation of IP prefixes, must be accurate for consumers. RFC 8805 [4] outlines factors for authenticating geofeed data, specifying procedures for validation before consuming self-published geofeed data. RFC 9092 [5] defines detailed authentication procedures, requiring the publisher to use Cryptographic Message Syntax (CMS) to create detached signature. The consumer must validate the signature, ensuring:

- The signer's certificate is part of the current manifest and covered by the Resource Public Key Infrastructure (RPKI) certificate [5].
- Path validation is carried out by using RPKI repositories (failure results in invalidation) — *Problem 1*
- The signed data is validated using the signer's public key certificate.
- The IP Address Delegation Certificate Extension [7] covers all address ranges in the geofeed file (failure to match results in invalidation) — *Problem 2*

However, based on our findings, these requirements are too rigid due to:

- Mismatches between the RPKI certificate resources and the current manifest.
- Limited compliance with certificate path validation, as only one AS adheres to this requirement mentioned in *Problem 1*.
- Many ASes combine data into shared files, conflicting with the requirement that geofeed file must cover all ranges as mentioned in *Problem 2*.

These shortcomings highlight the need for better integrity, ownership, and accuracy verification in geofeed data. To address this, we propose a new method to geofeed authentication, described in the next subsection.

A. Two-Step Approach

In this section, we describe our two-step approach to authenticate geofeed data. The first step involves authenticating the publisher of the geofeed, while the second step focuses on authenticating the geofeed data itself. By using this approach, we achieve the following objectives:

- Publisher Authetication: We verify the geofeed information for only a subset of prefixes owned by an AS, while also enabling the verification of multiple publishers to a shared file.
- Data Integrity and Trustworthiness: We ensure data integrity, determine authoritativeness, and establish non-repudiation for the published geofeed data.

1) Authenticating the Publisher: To authenticate the publisher of geofeed data, we propose leveraging a Public Key Infrastructure (PKI). Unlike the approach in [5], we rely on non-RPKI based repositories, enhancing the flexibility of the system for validation and verification. The PKI system binds the public key to the identity of the owner through the issuance of a certificate, ensuring the authenticity of the publisher.

Consumers can verify the authenticity of a geofeed publisher by tracing its certificate to a trusted Certificate Authority (CA), like Verisign. Once the CA confirms the certificate's validity, the consumer can trust that the publisher is who they claim to be, ensuring identity assurance.

In our experiments, we simulated the PKI ecosystem by generating over 1,800 unique certificates, each tied to an AS publishing geofeed data. For example, in Figure 4, we show three unqiue entities: a consumer, ISPs (ex: LS Networks, AT&T), and a Certificate Authority / Registration Authority (CA/RA) (ex: Verisign). LS Networks publishes a geofeed, signs it with its private key, and attaches a certificate. The consumer can then verify the certificate using the CA's public key, confirming the publisher's identity. This method mirrors the successful use of PKI in web authentication and can be easily adapted to geofeed validation, providing a reliable means to verify the authenticity of geofeed data.

In Fig. 4, the solid line with an arrow represents a request to fetch a signed geofeed file from an entity higher in the hierarchy, such as a consumer requesting a geofeed data

from LS Networks, and LS Networks requesting a signed geofeed file from AT&T, and so on. The dotted magenta line indicates the consumer accessing signed geofeed data from various entities, such as ISPs or RIRs. This process ensures the integrity of data and validates the publisher's identity, with each signature providing a layer of trustworthiness for the geolocation information.
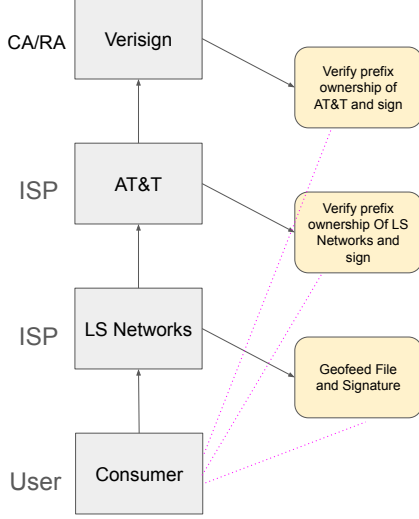
Prefix	RIPE		ARIN	
Comparison	RPKI Repository	ipinfo.io	RPKI Repository	ipinfo.io
Correct/Match	1109	6755	97	1104
Incorrect	1175	221	134	209
Missing	12244	7552	1294	212
Total	14582		1525	

TABLE II

PREFIX OWNERSHIP COMPARISON FOR RIPE AND ARIN BASED GEOFEED DATA.

Fig. 4. Authenticating Publisher and Geofeed Data through Digital Signatures

2) Authenticating Geofeed Data: Geofeed data is generally stable, with IP prefix geolocations rarely changing. This stability supports iterative signatures, ensuring only authorized entities can publish geofeed information.

For example, in Fig. 4, LS Networks signs a geofeed file, and consumers can verify this signature using LS Networks' previously validated public key. If the consumer trusts LS Networks, they can use the geofeed data directly. If further verification is needed, a larger ISP like AT&T can confirm LS Networks' authority over the prefixes, either through internal records or RPKI repositories. Once verified, AT&T can sign the geofeed file or LS Networks' signature, and the consumer can verify it with AT&T's public key. If the consumer trusts AT&T, they can accept the geofeed data.

If doubts persist, the process can continue further up the chain. For instance, ARIN (an RIR) may hold a prefix like 120.0.0.0/8 and lease it to AT&T, asserting AT&T's ownership of the prefix. AT&T, in turn, leases a sub-prefix (e.g., 120.1.1.0/24) to LS Networks, which can then validate LS Networks' geofeed data for that specific prefix. A CA like Verisign can add another layer of validation by confirming AT&T's prefix ownership and signing the data. This cascading verification process ensures that geofeed data is trusted at each level, with entities like ARIN asserting prefix ownership, and ISPs like AT&T and LS Networks providing further validation.

3) Evaluation of proposed approach: To assess prefix ownership, we compared data from RIR databases with two secondary sources: RPKI repositories and ipinfo.io (see Table II for RIPE and ARIN). We found low match rates in RPKI

repositories (RIPE at 7.6%, ARIN at 6.4%), reflecting RPKI's early adoption and limited coverage [8]. However, match rates were much higher when compared to ipinfo.io (e.g., RIPE at 46.3%, ARIN at 72.4%). This suggests that RPKI has room for improvement in coverage and adoption, while external datasets like ipinfo.io can supplement RPKI for more effective prefix verification. Simulating the signing and validation process for over 1,800 unique certificates from ASes publishing geofeed data, we observed higher match rates with ipinfo.io, demonstrating the flexibility and comprehensiveness of our approach, particularly for leased or reassigned prefixes.

Whereas RPKI focuses on IP address ownership and ROAs (Route Origin Authorizations) for prefix validation, our method authenticates geofeed data through a PKI-based chain of trust, verifying both prefix ownership and the publisher's identity. Unlike RPKI's emphasis on routing, our approach ensures the authenticity of geolocation data, which is crucial for decision-making. We eliminate the need for ROAs, as publishers directly sign geofeed data. This method scales well, since geofeed data changes infrequently, and requires only occasional updates to signatures. Validation involves checking certificates and signatures in the trust chain, providing flexibility compared to RPKI's centralized prefix validation.

We acknowledge the study's limitation due to the lack of real-world data for testing our proposed authorization method, which, while simple and effective in theory, requires validation in practical settings.

IV. CONCLUSIONS AND FUTURE WORK

In conclusion, our analysis of geofeed data shows that RIPE lead in adoption among RIRs, with ASes in the Business category driving geofeed use. However, publishers do not fully adhere to proposed standards, highlighting a gap between recommended practices and real-world implementation. To address these challenges and enhance the validation of geofeed data, we propose a multi-step approach for authentication.

Future work includes real world testing to assess performance and exlore deloyment challenges. We will also describe the detailed steps for preparing geofeeds for analysis, guiding future implementation. A key next step is using multiple geofeed snapshots over extended periods to study geofeed dynamics. Simplifying authentication requirements would help consumers leverage existing infrastructure, while providing publishers more flexibility to meet standards. Incorporating multiple secondary sources of prefix ownership will further refine and enhance our approach to geofeed authentication.

REFERENCES

[1] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-Based Geolocation of Internet Hosts," *IEEE/ACM Transactions on Networking*, vol. 14, pp. 1219–1232, Dec. 2006. Conference Name: IEEE/ACM Transactions on Networking.

[2] S. Ding, X. Luo, J. Wang, and X. Fu, "Gnn-geo: A graph neural network-based fine-grained ip geolocation framework," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 6, pp. 3543–3560, 2023.

[3] O. Dan, V. Parikh, and B. D. Davison, "Ip geolocation through geographic clicks," *ACM Trans. Spatial Algorithms Syst.*, vol. 8, Mar. 2022.

[4] E. Kline, K. Duleba, Z. Szamonek, S. Moser, and W. A. Kumari, "A Format for Self-Published IP Geolocation Feeds," Request for Comments RFC 8805, Internet Engineering Task Force, Aug. 2020.

[5] R. Bush, M. Candela, W. A. Kumari, and R. Housley, "Finding and Using Geofeed Data," Request for Comments RFC 9092, Internet Engineering Task Force, July 2021.

[6] S. Murphy, C. Villamizar, C. Alaettinoglu, and D. Meyer, "Routing Policy System Security," Request for Comments RFC 2725, Internet Engineering Task Force, Dec. 1999.

[7] D. C. W. L. Jr., K. Seo, and S. Kent, "X.509 Extensions for IP Addresses and AS Identifiers." RFC 3779, June 2004.

[8] Ryan Polk, MANRS, "The united states government can take lead in rpki deployment," 2024.

[9] "Developed Countries 2024 — worldpopulationreview.com." https://worldpopulationreview.com/country-rankings/developed-countries. [Accessed 10-10-2024].

[10] "Developing Countries 2024 — worldpopulationreview.com." https://worldpopulationreview.com/country-rankings/developing-countries. [Accessed 10-10-2024].

[11] R. Roberts, G. Huston, and D. T. Narten, "IPv6 Address Assignment to End Sites." RFC 6177, Mar. 2011.

APPENDIX

A. Ethics

This work deals with only publicly available data and does not raise any ethical issues.

B. Data Visualization

To evaluate the gathered data, we visualize geofeed in two ways: a world map and a heatmap. The first way is to have a world atlas that shows color gradient in a country-level, based on the IP prefix counts from `inetnum` and `inet6num`. The second way is to have a plot where the x-axis is country and the y-axis is IP prefix from `inetnum` and `inet6num`. The original plots can be found in our GitHub. The following are the four research questions that we come up with.

1) World Map: do developed countries report geofeed more than developing countries?: Our hypothesis in the World Map is that the published geofeed is predominantly from developed countries [9], [10]

We use `inetnum` and `inet6num` records from all five RIRs, namely AFRINIC, APNIC, ARIN, LACNIC and RIPE NCC. For the World Map, we also use GIS (Geographic Information System) data from ARCGIS.com to plot in a world map format. We join those datasets together on ISO 3166 country codes.

As seen from Figure 5, our hypothesis is not valid. The published geofeed are mostly from developed countries, but not necessarily; they are darker in the color gradient. The country with the most IP prefix count in each RIR is Germany in AFRNIC, Thailand in APNIC, US in ARIN, Argentina in LACNIC, and Russia in RIPE. Most of the empty countries are from undeveloped countries; they are hatched with red lines.

It is interesting that the only a few countries from Africa have reported geofeed, even in AFRINIC. Developed countries outside of Africa such as Germany, Russia, United States, and some European countries reported more than any country from Africa. This discovery still validate our hypothesis, yet this can be a further research topic.

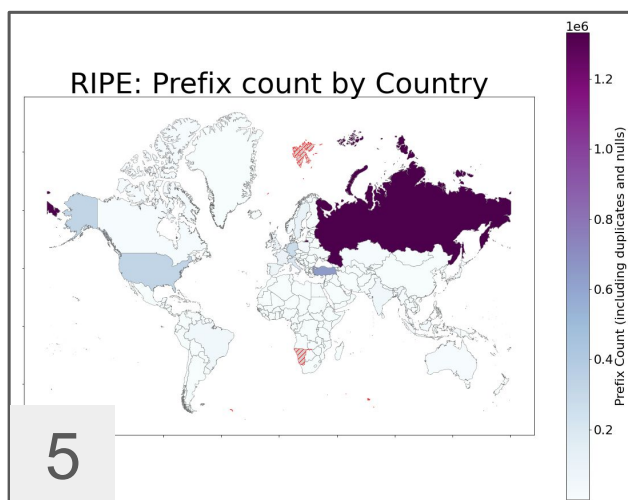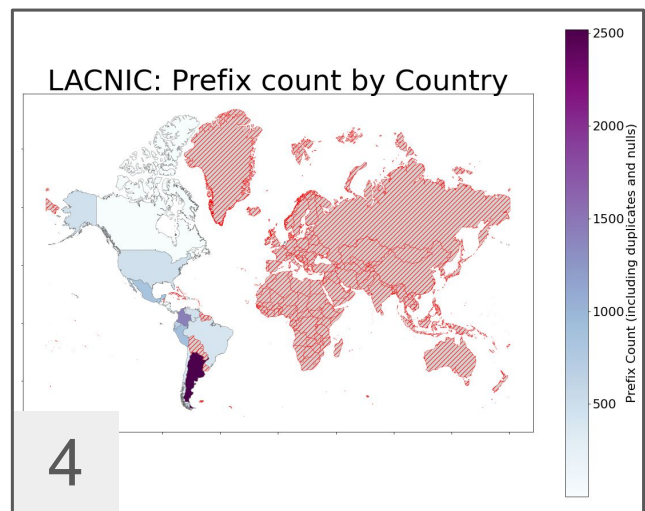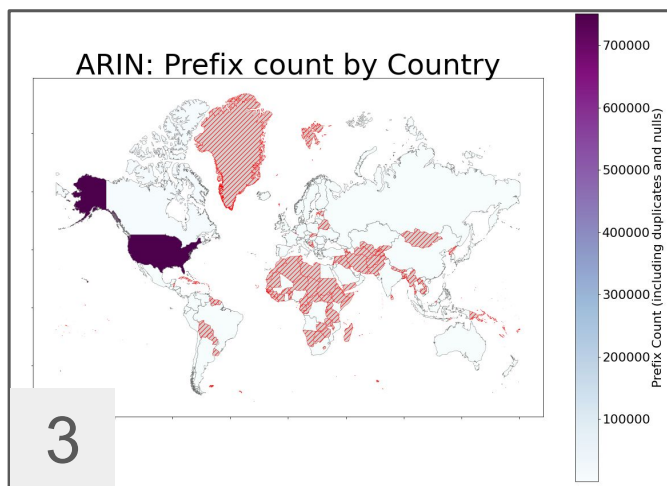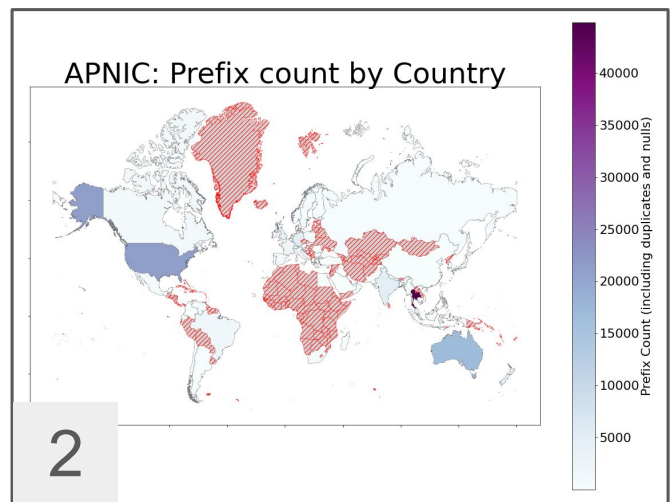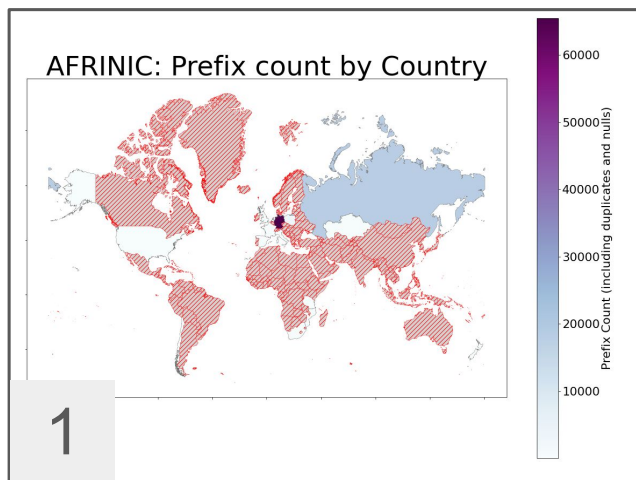*2) Heatmap:* We use heatmap to understand geofeed in terms of the two IP address systems: IPv4 and IPv6.

As a methdology, we use `inetnum` and `inet6num` records from all five RIRs, namely AFRINIC, APNIC, ARIN, LACNIC and RIPE NCC. We split them into IPv4 and IPv6. We count the number of IP prefix by prefix length. Particularly for IPv6, we only show prefix lengths that are multiples of 4. Due to the skewed distribution over countries, we only show countries with at least 5% portion of the overall IP prefix count in the relevant RIR.

3) IPv4: is the commonly used IPv4 prefix /24 also common in geofeed?: Our hypothesis is that /24 is the most common IPv4 prefix in geofeed. /24 is indeed the most commonly used IPv4 prefix, because this provides a balance between the number of available hosts and efficient use of IP address space, making it suitable for small to medium-sized networks.

As seen from Figure 6, our hypothesis for IPv4 is valid. All RIRs but LACNIC show that /24 has the most number of IP prefix count. The country with the most IP prefix count in each RIR is Germany in /24 in AFRNIC, Thailand in /24 in APNIC, US in /24 in ARIN, Argentina in /22 in LACNIC, and Russia in /24 in RIPE.

4) IPv6: is the recommended IPv6 prefix /64 common in geofeed?: Our hypothesis is that /64 is the most common IPv6 prefix in geofeed, because IETF (Internet Engineering Task Force) recommends that by being selective among prefixes /48, /64, and /128 from a previously announced RFC [11].
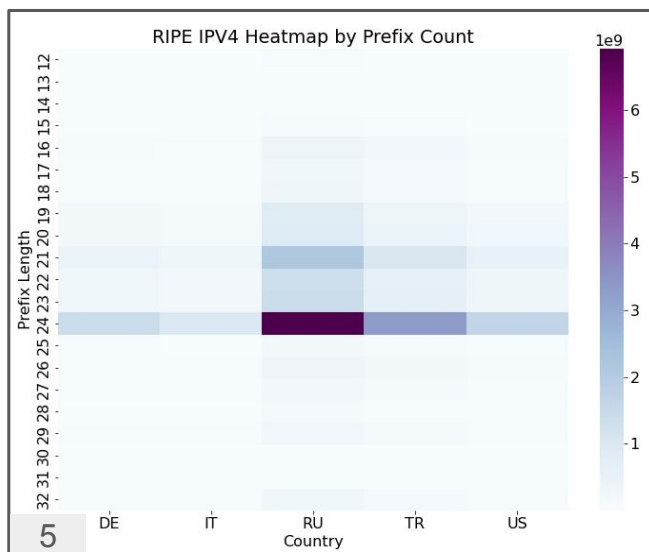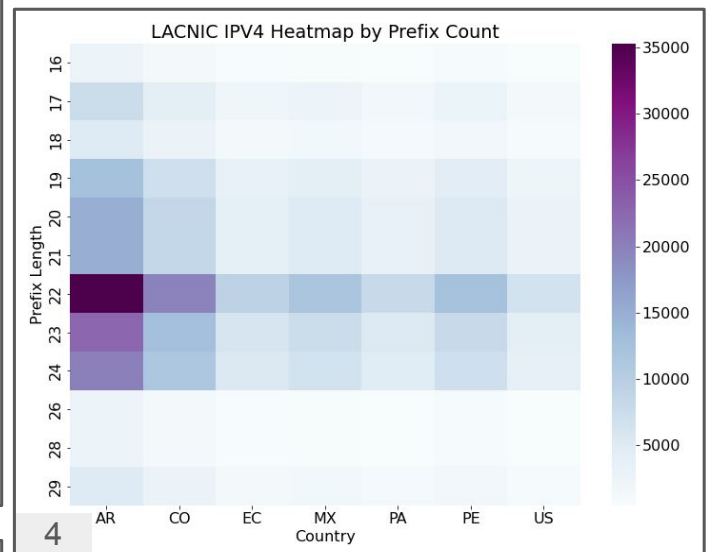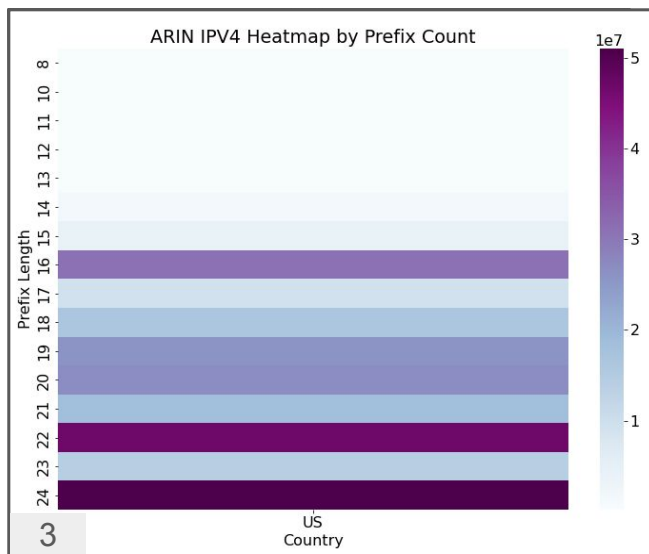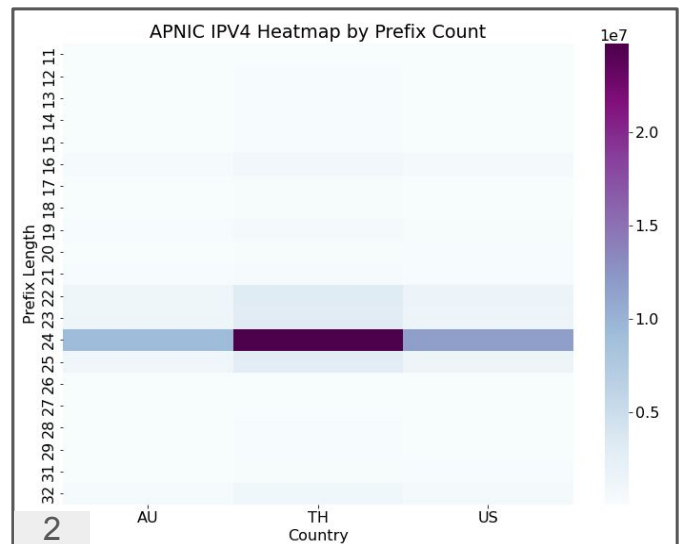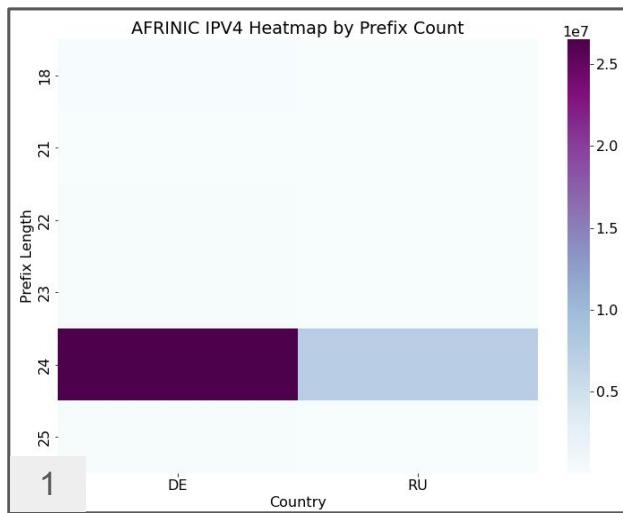
As seen from Figure 7, our hypothesis for IPv6 is not valid. It turns out that /32 is the most commonly reported prefix in IPv6. The country with the most IP prefix count in each RIR is Germany in /36 in AFRNIC, Thailand in /48 in APNIC, US in /32 in ARIN, Argentina in /32 in LACNIC, and Russia in /32 in RIPE.

The darker, the more
IP prefix counts.

Fig. 5. World Map by Prefix Count.
The country with the most IP prefix count in each RIR is 1: Germany in AFRNIC, 2: Thailand in APNIC, 3: US in ARIN, 4: Argentina in LACNIC, and 5: Russia in RIPE.

The darker, the more IP prefix counts.
x-axis: country
y-axis: IPv4 prefix

Fig. 6. IPv4 Heatmap by Prefix Count.
The country with the most IP prefix count in each RIR is 1: Germany in /24 in AFRNIC, 2: Thailand in /24 in APNIC, 3: US in /24 in ARIN, 4: Argentina in /22 in LACNIC, and 5: Russia in /24 in RIPE.
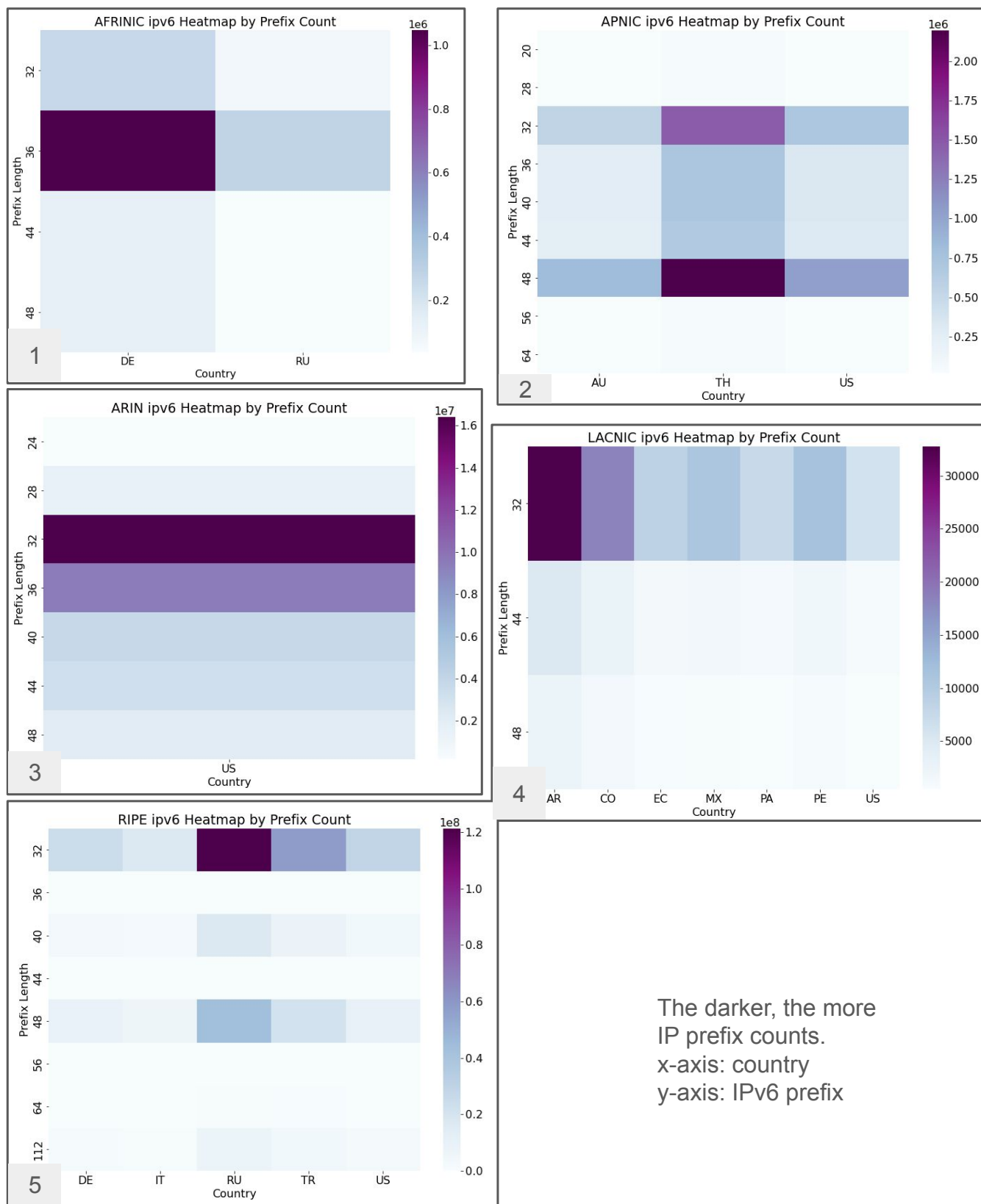
Fig. 7. IPv6 Heatmap by Prefix Count.
The country with the most IP prefix count in each RIR is 1: Germany in /36 in AFRNIC, 2: Thailand in /48 in APNIC, 3: US in /32 in ARIN, 4: Argentina in /32 in LACNIC, and 5: Russia in /32 in RIPE.