Enabling a longitudinal study of Geofeed data quality, adoption, and dynamics

Ricky Mok, Parwiz Khan (CAIDA)
Eduard Dehelean, Rocky Chang (Calvin University)

This work is partially supported by NSF #2212241.

A 20+ years research problem

An attempt to solve a long-standing (and ever completely solved) Internet research problem:

- IP geolocation!

Existing solutions have their limitations

- IP geolocation databases (proprietary methodologies, unknown accuracy)
- RTT triangulation (Vantage point distribution, non-responsive IPs)
- DNS hostnames (unconventional/ambiguous abbreviations, stale DNS records)

What is Geofeed?

One trivial solution: Let the operators tell you the answer :)

RFC 8805 and RFC 9632 were published in 2020 and 2024, respectively, standardized the format and the method for network operators to publish IP geolocation data

APNIC WHOIS

```
inetnum:    134.159.0.0 - 134.159.255.255
netname:    TELSTRAGLOBAL-AP
descr:      Telstra Global Internet Services Network Blocks
country:    AU
…
geofeed:    https://geofeed.telstraglobal.net/geofeed.csv
```

Geofeeed.csv

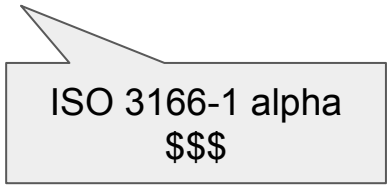
```
134.159.243.0/24,AE,,,
…
134.159.7.0/24,AU,,,
134.159.40.0/24,AU,,,
134.159.41.0/24,AU,,,
134.159.54.0/24,AU,,,
134.159.92.0/24,AU,,,
134.159.93.0/24,AU,,,
134.159.94.0/24,AU,,,
…
```

Geofeed

Use WHOIS attribute `geofeed:` and/or `remarks: Geofeed` to link to a CSV file as “geofeed” [RFC9632]

Each line in the CSV file, namely geofeed entry, contains geolocation information of a prefix

- `ip_prefix,alpha2code,region,city,postal_code`



Everything sound good. But…

Only `ip_prefix` field is required. Others are optional. [RFC8805]

- Geofeed could have no geo-information.

Human factor is inevitable

- Non-compliant format. Google doc rather than CSV
- Fat finger. We found an IPv6 geofeed entry has a prefix length of /2
- Stale records.

Prior work studying geofeed

Focus on coverage and consistency with other IP geolocation data

- Rahel A. Fainchtein and Micah Sherr. You can find me here: A study of the early adoption of geofeeds. In Passive and Active Measurement Conference (PAM), 2024.
- Ioana Livadariu, Kevin Vermeulen, Maxime Mouchet, and Vasilis Giotsas. Geofeeds: Revolutionizing IP geolocation or illusionary promises? In ACM CoNEXT, 2024.

Our study

Longitudinal study on the geofeed data quality

- Disseminate in the correct ways
- Format compliance

X Comment on the “accuracy” of the geolocation information published by the feeds

Data Pipeline

Datasets:

https://catalog.caida.org/dataset/geofeed_whois

Do the operators publish Geofeeds in a correct way?

We identify (potential) Geofeed by applying string match “geofeed” to the entire WHOIS dump.

- Cover not only WHOIS attribute `geofeed:` and `remarks: Geofeed`
- But also capture Geofeed published in “non-standard” way

Non-standard geofeed publication

Some number of geofeeds were not published using the two dedicated WHOIS attributes. E.g., using desc:

All geofeeds published in ARIN used `Comment: Geofeed`

RIR	Types		
	Std	Nonstd	Total
RIPE	2317	3	2320 (69.4%)
ARIN	0	816	816 (19.6%)
APNIC	319	1	320 (9.1%)
LACNIC	21	0	21 (0.6%)
AFRINIC	48	0	48 (1.4%)
Total	2705 (76.7%)	820 (23.3%)	3525

Duplicated geofeed entries

We identified unexpectedly high number of duplicated geofeed entries published in different geofeeds

RIR	Before entry dedup	After entry dedup		
	Total	Unique	nonunique	Total
RIPE	480051 (94.5%)	369646	54258	423904 (94.3%)
ARIN	20952 (4.1%)	20323	542	20865 (4.6%)
APNIC	2453 (0.5%)	2030	302	2332 (0.52%)
LACNIC	4303 (0.85%)	40	2130	2170 (0.48%)
AFRINIC	55 (0.05%)	55	0	55 (0.1%)
Total	507814	392094	57232	449326

Work in progress

- Broken links
  - opengeofeed.org had outages in Nov 2025. Impacted >60 links.
  - singularity-telecom.com is no longer reachable. Impacted 77 links.
- Wrong format
  - Google spreadsheets, dropbox links, HTML…
- Geofeed “hijack”
  - Provide geofeeds for prefixes that the operators have no control
- Update frequency