RIPE IPmap - The RIPE NCC's Approach to Infrastructure IP Geolocation

Robert Kisteleki, RIPE NCC

Most contemporary IP geolocation services and approaches focus on **end-user ("eyeball") geolocation** - mainly because that's where the most understood business need is; payment process verification, geofencing, content localisation heavily depend on such services.

However, there's also a visibly growing need to **geolocate infrastructure components**. As of yet this is a much less supported case. Some motivations for such data sets include answering questions like:

- How did packets travel physically / geographically?
- Did packets go through one or more IXPs?
- How "local" is the traffic; does it cross country or continent boundaries, does it exhibit unexpected detours?
- Support analyses about network disruptions such as infrastructure issues, natural disasters, etc.

The RIPE NCC, as part of our information services portfolio, has created a prototype and later a preliminary service we today call "RIPE IPmap" to address this need. The intention is to use public data and measurement tools to compile an open data set containing geolocation information for IP infrastructure components - think of hops of a traceroute path.

Our technical approach is to use various methodologies (algorithms) to come up with a number of "best guesses" about where particular IPs are used around the globe at a particular point in time. These methodologies may be related, or may be completely independent from each other, but they should all provide their own assessment. Our conceptual collection of methodologies currently includes:

Short name	**Short description**	**Reliability**	**Difficulty**	**Absolute/ relative?**	**Iterative?**	**Notes**
Registry databases	RIPE DB, ARIN DB, APNIC DB, ...	Medium	Easy	Absolute	No	Available baseline, no precision guarantees
Other public databases	PeeringDB et al	High	Easy	Absolute	No	Well maintained for what they cover
DNS names	Use reverse DNS lookups, extract location hint based on predefined templates	Medium	Hard	Absolute	No	Discoverable, templatable

Geofeeds / RFC8805	Use data published in this format	High	Medium	Absolute	No	Authoritative but as of yet scarce
Crowdsourcing	Ask people with the know	Varies	Medium	Absolute	No	Low bar for entry, quality depends on "who says that"
Interface aliases	Research dataset about what size IP prefixes are assigned to interfaces of the same physical router	High	Easy	Relative	Yes	Research algorithm needs stable execution
"Triangulation"	Low RTT from points with known location	High	Hard	Absolute	No	Should start with a good ground truth
Proximity	RTT difference between trace hops mean they are close to each other	High	Hard	Relative	Yes	Signal can get stronger with increased amount of observations
Gap filling	Before and after in the same place, RTT difference "not too high"	Low	Medium	Relative	No	
Gamification	Asking people to guess	Low	Medium	Relative	N/A	Low scalability
Historical archives	Other (e.g. paper) documentation from the past	Low	Hard	Absolute	No	Good for legacy space?

(Notes: the table shows possibilities, as of yet only a few of these are implemented in our system. “Absolute” is where the method gives a possible location on its own, “relative” is when this is based on data already present in the system. “Iterative” methods extend the dataset by introspecting already existing data and extending those.)

Based on the outputs of these algorithms - and perhaps more in the future - an overarching **combiner algorithm** summarises the results. It compares the assessments of the various other methodologies, checks for consistency or disagreements, and outputs the conclusion (“best guess”) for the location of particular IP addresses. The output may contain multiple locations, with evaluated probabilities as well, e.g. “this IP is in location X with 70% probability and in location Y with 30% probability”. Consumers of the combiner’s output can decide to only use the best guess or the various alternatives as well.

The IPs to be considered for IP infrastructure geolocation usually come from hops in previously collected traceroute data, such as from RIPE Atlas. Using the published ultimate output, it should be possible to annotate ad-hoc traceroute outputs as well. There are a finite set of IP addresses involved in IP infrastructure; early estimations showed this number to be in the low millions (ballpark).