Measurement Vantage Point Selection Using Similarity Metrics

Team
XXX
XXX
XX@XX.XX

ABSTRACT

When measuring the Internet it matters what vantage points are used to conduct measurements. We propose a novel means of selecting vantage points, which is not based on categorical properties (like origin ASN, or geographical location), but is based on the topological (dis)similarity between vantage points.

We show the implementation of a similarity metric between RIPE Atlas probes and show how it performs better for the purpose of topology discovery, than the probe selection mechanism that is built into RIPE Atlas.

ACM Reference format:
Team. 2017. Measurement Vantage Point Selection Using Similarity Metrics. In *Proceedings of 2016 Applied Networking Research Workshop, Prague, Czech Republic, July 2017 (ANRW'17)*, 4 pages.
DOI: http://dx.doi.org/-----

1 INTRODUCTION

In a measurement system with a wide selection of vantage points, it can become a challenge to select the most appropriate vantage points for a given measurement. RIPE Atlas, does have over 9600 active measurement vantage points, which can be selected for measurement based on categorical properties, like origin ASN and country.

When doing a measurement, a user is limited in how many vantage points to use for his/her measurement. This is not only due to limitations that he measurement platform imposes, but also collecting data from a larger number of vantage points means a larger volume to analyse and store. So it makes sense to optimize for a minimal set of vantage points with a maximum chance of observing the phenomenon the user is interested in.

Network operators may need to debug a network service with only limited information about the problem ("our network is slow for users in France!"), so doing a minimal set of measurements that would allow selecting a wide diversity of networks could be a valuable add-on to the tools available to network operators. If one can say "give me 10, as diverse as possible vantage points in France", instead of hand-picking 10 vantage points, based on the current primitives, which are physical location and origin ASN. If a user does have a categorical specification (like "France"), a diversity metric would allow to select the most dissimilar vantage points, in an attempt to explore a networking phenomenon from as diverse angles as possible.

If one finds an interesting networking phenomenon, one could now again use the similarity metric to it's advantage by selecting the most similar vantage points to the one exhibiting the phenomenon, in an attempt to validate the phenomenon not being a vantage point specific artifact.

ANRW'17, Prague, Czech Republic

———-…$15.00
DOI: http://dx.doi.org/-----

2 SIMILARITY METRIC

2.1 Topological Similarity

We aim to quantify the topological similarity (or diversity) of RIPE Atlas. We say that two probes are topologically similar if the relative network distance (in terms of routing hops) separating them is small. We argue that topological distance is more relevant from a measurement point of view as it is directly based on network data rather than traditional line of sight distance, which are network-agnostic.

Measuring topological similarity is useful as it can help in probe selection (e.g., by warning the user that the probe she is using are similar or indeed, diverse) and it can help in future deployment of the infrastructure (e.g., where to deploy new probes).

While it is more useful, topological distance is also harder to capture in practice as it depends on several dimensions. For instance, probes in the same AS can actually see very different paths (e.g., if there are connected to different egress routers), while probes in different AS can see similar paths (e.g., if there are connected close to the same IXP and the ASes have similar routing policies).

In the following, we show how we can capture these nuances in a single metric. It is based on the Jaccard similarity coefficient. It captures the expected similarity between two probes for the paths towards any single destination. Users interested in monitoring their ASs from multiple vantage points at a small measurement cost may use it to compute a set of most diverse probes according to that metric.

We present the results of computing the this metric for all pairs of RIPE Atlas probes.

We show that it successfully captures network proximity with probes in the same AS being significantly more similar than probes in different ASes.

2.2 Measuring topological similarity

We now define our our measurement-based metric and discuss the practical relevances.

Jaccard-based similarity Let x and y be the two probes, we define M as the set of measurements performed by both x and y in the interval t. Traceroutes in M let us discover a set of IPs discovered from x and another set of IPs crossed from y. We note these sets P_x

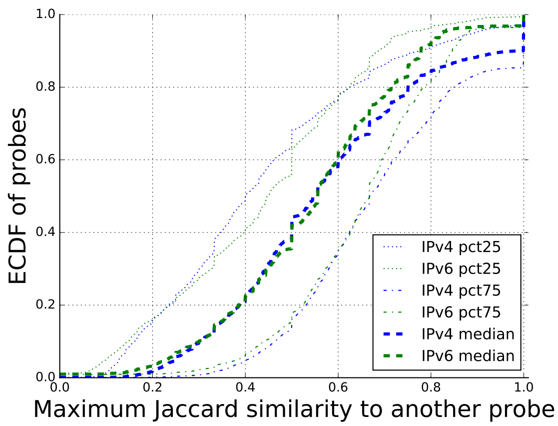

Figure 1: About 10% (resp. 5%) of the RIPE Atlas probes see a median Jaccard index of 1.0 in IPv4 (resp. IPv6). The large interquartile range results from the variability in the number of unique IPs encountered by each measurement.

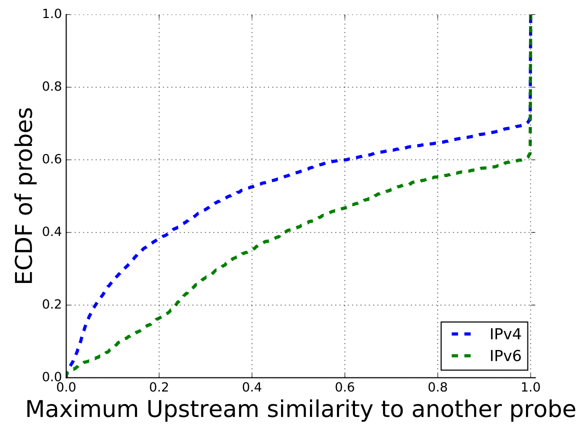

Figure 2: Being more coarse-grained, the upstream metric classifies more probe as similar, with up to 30% (resp. 40%) of the pairs being considered as strictly similar in IPv4 (resp. IPv6). The distribution is almost binary, probes tend to be very similar or not at all.

and P_y, respectively. For each destination as defined in RIPE Atlas measurement specifications (i.e. an IP or hostname destination) we calculate the similarity of results.

Based on this, we define the Jaccard similarity coefficient of x and y as: $d = \frac{|P_x \cap P_x|}{|P_x \cup P_x|}$.

Intuitively, a result of 1 for a pair of probes indicates they discover the same set of IP adresses for at least half of the measurement specifications, and a result close to 0 indicates very few IP addresses were the same for at last half of the measurement specifications that were in common between the two probes.

To compute the coefficient in practice, we rely on traceroute results for a single day to compute d. For each pair of probes we find the measurement specifications these probes have in common, this is typically tens of specifications, and at least includes all built-in measurements (§??). For each destination that two probes have in common, we calculate the ratio of the number of IPs (not considering private addresses) seen by both probes over the union of the IPs discovered by traceroutes from the two probes. This way we end up with a list of Jaccard indexes per probe pair. To ensure statistical significance, we only consider the Jaccard indexes if the list is long enough ($n \geq 17$) and consider the 25th, 50th and 75th percentile of the lists as actual coefficients. Doing so makes the metric resistant to outliers even though the metric is sensitive to the set of common destinations between two probes. If the list is not long enough, we leave the Jaccard metric for the probe pair undefined. We calculate the metric for IPv4 and IPv6 separately, since these topologies are not congruent.

Practical relevance The metric depends on the actual measurement performed by the probes which bears the question of their practical relevance. As discussed in §??, every single probes perform a number of (identical) built-in measurements, ensuring a base level of common measurements. Consequently, we were able to compute the Jaccard metric in IPv4 for 9318 probes.

2.3 How topologically similar is RIPE atlas?

We now compute the metric for all the pairs of probe considering one day (31 March 2016) worth of traceroute measurements. We show that about 30% (resp. 10%) of probes are similar when considering the upstream (resp. Jaccard-index). We also show the added value of both two metrics with respect to simple distance-based metric. Indeed, even though geographically close probes tend to exhibit higher metric values, our metrics manage to capture additional interesting topological properties (such as similarity over long-distances) or simply incorrect geolocation data for the probes.

Jaccard-based similarity Figure 1 plots the CDF of probes with regards to the maximum similarity to another probe, considering the 25th, 50th and 75th percentiles of the Jaccard metric (see §2.2). We see that about 10% (resp. 5%) of probes see a median Jaccard index of 1.0 in IPv4 (resp. IPv6). We also see that interquartile range is relatively large. This is due to the fact that there is understandably a large variability in the fraction of unique IPs encountered between measurement specifications for a given probe pair.

Upstream-based similarity Figure 2 plots the same CDF as above consider the upstream-based similarity instead. We see that the upstream metric classifies strictly more probes as similar with close to 30% (resp. 40%) of the probes seeing a metric of 1 in IPv4 (resp. IPv6). Intuitively, we can explain this as the Jaccard index considers all the IPs address on the path in its computation while the upstream metric only considers one. It therefore tends to be higher. Another observation is that the distribution of the upstream-based similarity is almost binary. Each pair of probes tend to see a high or a low value. As such, the metric acts can act as a binary filter.

Validating the metric against physical distance and detecting geolocation discrepancies Intuitively, we expect probes pair with a result close to 1 to be physically close. Table 1 shows how physical distance and being part of the same Autonomous System or not affect the median Jaccard-index. We compute the distance between

	IPv4	IPv6
Probe pairs in the same AS		
# with Jaccard ind. ≥ 0.9	1805	56
25th percentile distance	7 km	0 km
50th percentile distance	40 km	2 km
75th percentile distance	104 km	17 km
Maximum distance	8,817 km	664 km
Probe pairs in different ASes		
# with Jaccard ind. ≥ 0.9	11	0
25th percentile distance	2 km	-
50th percentile distance	9 km	-
75th percentile distance	112 km	-
Maximum distance	532 km	-

Table 1: Pairs of probe with high Jaccard coefficient (≥ 0.90) tend to be geographically close (according to the location given by the probe owner) to each other. The maximum distance (8,817km) turned out to be a geolocation error than the metric helped catching.

	IPv4	IPv6
Probe pairs in the same AS		
# with IP upstream ind. ≥ 0.9	4617	5065
25th percentile distance	8 km	50 km
50th percentile distance	32 km	175 km
75th percentile distance	99 km	504 km
Maximum distance	8817 km	13753 km
Probe pairs in different ASes		
# with IP upstream ind. ≥ 0.9	191	73
25th percentile distance	7 km	108 km
50th percentile distance	124 km	345 km
75th percentile distance	288 km	1076 km
Maximum distance	6288 km	7636 km

Table 2: Similarly, pairs of probe with high upstream metric value (≥ 0.90) tend to be geographically close. With respect to Jaccard, the upstream metric classifies more pairs as similar as it is more coarse-grained.

probes based on the geographical coordinates provided by the probe host.

As expected, the similarity metric is higher when the corresponding probes pair belong to the same AS. Yet, we observe that few probes in different ASs produce similar results according to the Jaccard metric. In IPv4, 1,816 pairs of probes have a Jaccard metric higher than 90%, while 90.4% of these pairs are separated by less than 200 km. Geographically far apart probes that have a high Jaccard metric could indicate interesting topologies (e.g. long-distance tunneling) or incorrect geolocation data. Indeed, our manual inspection of few geographically far apart probes with a high Jaccard metric revealed a few probe geolocation errors. When we contacted the probe hosters, we found they forgot to update probe location after it moved, which was then corrected. In general, we believe that the relatively high correlation between geographic location and the Jaccard index can be used as a tool to easily spot probe geolocation errors, and so produce a better platform, for purposes where precise geolocation is of utmost importance. Table 2 shows the corresponding results with the upstream metric. Again, with physical distance the likelihood of probe pairs being similar decreases, and more often probe pairs in the same ASN are more similar, which matches expectation.

3 EXPLOITING SIMILARITY

3.1 Exploiting probe similarity

Ensure a better measurement continuity RIPE Atlas probes typically have some downtime. For example, as of May 11, 2016, 4,421 probes (32%) were disconnected. To reduce the chances for gaps in a time series of measurement results, users could run their measurements on multiple similar probes instead of only one. In case one of the probes disconnects, the user would still have its measurement running on the backup probe(s).

Help diagnostics on the path up to the common upstream IP Topological similarities can also help diagnostics on the path up to and including the common upstream IP. Especially if the common IP is the first public IP address in the IP path, this allows for better diagnostics of physical infrastructure between probe and the public Internet.

Improve users capabilities Combining similar probes could help users to improve their capabilities. For example, it could help users to bypass the limitations imposed by RIPE Atlas. To increase measurement frequency (initially limited to one measure per minute), users can launch the same measurement on similar probes. By combining the results of two probes, users can get one result every 30 seconds instead of one every minute.

Improve measurements precision When launching a large set of measurements on only one probe, users may overload a probe and lose precision [**?**]. Dividing a set of measurements between similar probes could give users the ability to run a large number of measurements without losing precision. Similarly, if one probe is loaded by someone else, users could use a similar but not loaded probe so as to improve its precision.

Boost IP topology discovery by 25% An additional way to exploit our probe similarity metrics is to (dis)cover as much of the IP topology address space as possible, given a limited probing budget. As the Jaccard-based similarity metric is designed to capture (dis)-similarity between probes based on IP topology, we explored exploiting this by performing the following experiment: We ran 1002 experiments in which we compared RIPE Atlas' probe selection [**?**] to a custom RIPE Atlas probe selection mechanism that selects probes based on their maximum dissimilarity. To prepare for an experiment we select a random probe *pr_init* and one of the three commonly used probe groupings (*asn_v4*,*asn_v6* and *country_code*). If there are 10 or more probes available in the same group as *pr_init*, we run an experiment that compares RIPE Atlas' default selection mechanism to a maximum dissimilarity based selection mechanism: First we select a probe count (*n_pr*) randomly between 2 and the minimum of 100 and 1/3 of available probes in that group, and a random destination d from the set of .1 addresses for all globally routed IP prefixes. We create a set of maximally dissimilar probes $Sdis$, starting from *pr_init*, by adding the probe that has the lowest Jaccard similarity score to any of the probes in the set, until we reach the required count *n_pr*. We perform a set of traceroutes from $Sdis$ to d as well as a reference set of traceroutes by making RIPE

Atlas select *n_pr* probes from the same group and traceroute to *d* from these. We compared the total number of IP addresses discovered normalized by number of probes that actually performed measurements (this can be less than specified). We found that our maximally dissimilar probe selections, on median discovered 25% more IP addresses per probe than RIPE Atlas' probe selection with the same parameters.

4 CONCLUSIONS

Outcomes:

Data: provide a page with links to all of the measurement data!

REFERENCES