Improving AS Relationship Inference Using PoPs

Lior Neudorfer
Tel-Aviv University
Email: liorus@gmail.com

Yuval Shavitt
Tel-Aviv University
Email: shavitt@eng.tau.ac.il

Noa Zilberman
Tel-Aviv University
Email: noa@eng.tau.ac.il

***Abstract*— The Internet is a complex network, comprised of thousands of interconnected Autonomous Systems. Considerable research is done in order to infer the undisclosed commercial relationships between ASes. These relationships, which have been commonly classified to four distinct Type of Relationships (ToRs), dictate the routing policies between ASes. These policies are a crucial part in understanding the Internet's traffic and behavior patterns. This work leverages Internet Point of Presence (PoP) level maps to improve AS ToR inference. We propose a method which uses PoP level maps to find complex AS relationships and detect anomalies on the AS relationship level. We present experimental results of using the method on ToR reported by CAIDA and report several types of anomalies and errors. The results demonstrate the benefits of using PoP level maps for ToR inference, requiring considerable less resources than other methods theoretically capable of detecting similar phenomena.**

I. Introduction

Inferring the commercial relationships between service providers is an important line of research. The knowledge gained through the understanding of commercial relationships is used in research on Internet routing, can improve network performance as well as help increase its robustness. However, commercial relations between service providers are interesting first and foremost as they determine BGP routing policies between ASes. Contractual commercial agreements between Administrative Domains (which control Autonomous Systems) are usually not publicly disclosed, as so inferring them from measurement data has been a focus of many works. These relationships can be classified into three Types of Relationships (ToR) [1]: customer-to-provider (c2p), peer-to-peer (p2p), and sibling-to-sibling (s2s). Gao [2] was the first to present a method of inferring these relationships from publicly available BGP route data, and introduced the **valley free** AS path rule. An AS path is considered **valley free** if it consists of an uphill segment (customer to provider links), followed by an optional peer to peer link and a downhill segment (provider to customer links). Subramanian *et al.* [3] formally defined the "ToR Problem" as an optimization problem that seeks to find a ToR labeling for an AS graph which maximizes the number of valley-free paths. Di Battista *et al.* [4] and Erlebach *et al.* [5] showed that the ToR problem is NP-complete, and developed mathematically rigorous approximate solutions to the problem. Dimitropoulos *et al.* [6] acknowledged that a solution that maximizes the number of valley-free paths is not necessarily correct, and improved AS relationship detection by taking AS degrees into consideration. Shavitt *et al.* [7] suggested a near-deterministic algorithm for solving the ToR problem using an Internet Core, a subgraph of the Internet graph which contains the top-level providers. Their algorithm inferred AS relationships in AS paths by examining their relation to the Internet core.

The relationship between two ASes is sometimes more complex than a single ToR between all border routers. Gao [2] mentioned complex AS relationships as a cause for excessive sibling-sibling ToR inference. Subramanian *et al.* [3] introduced AS path anomalies as specific patterns which cause paths not to be valley free. Dimitropoulos *et al.* [6] conducted a survey with several large ISPs, and revealed backup links and hybrid c2p/p2p relationships. A hybrid relationship is one in which two ASes connect in multiple peering points and have different types of relationships at these points.

There are several levels the Internet maps are presented at, each level of abstraction is suitable for studying different aspects of the network. The most detailed level is the IP level, while the most coarse level is the Autonomous System (AS) level. An interim level of aggregation between the router and the AS level graphs is the PoP level. A PoP is a group of routers which belong to a single AS and are physically located at the same building or campus. PoP level maps have been constructed from various data sources. Andersen *et al.* [8] used BGP messages for clustering IPs and validated their PoP extraction based on DNS. Rocketfuel [9] generated PoP maps using tracers and DNS names. The iPlane project [10] generated PoP level maps by first clustering IP interfaces into routers by resolving aliases, and then clustering routers into PoPs by probing each router from a large number of vantage points and assuming that the reverse path length of routers in the same PoP will be similar. The DIMES project, takes a structural approach and looks for bi-partite subgraphs with certain weight constraints in the IP interface graph of an AS [11]. The bi-partites serve as cores of the PoPs and are extended with other nearby interfaces.

This paper proposes a method that accepts as an input a collection of traceroutes and IP to PoP mapping, converts the traceroutes to PoP level traceroutes, and analyzes the ToR at the PoP level. The analysis at this level reveals oddities that help us make several contributions, which can be roughly classified into two classes. First, by looking at Valley-freedom violation we can easily detect imperfections in our data-set inputs: errors in the initial ToR assignment, missing sibling relationships, missing IXP address prefixes, and erroneous IP to AS mapping. Second, using the same method we can identify complex ToRs, a holy grail in the

field. An interesting subgroup of complex ToRs we identified are "academic oddities": cases where academic networks do not follow the strict commercial rules of relationships. While some of our findings can be achieved at the IP level, we point out that the analysis at the PoP level dramatically reduces the processing amount.

II. Analysis Process

As said above, this paper proposes a method for inferring ToRs at the PoP level, and for discovering anomalies that lead to improvements of its input data (ToRs, IP to AS mapping) as well as revealing complex AS relationships. We start by converting a traceroute dataset to a PoP level traceroute (preprocessing); then we deduce missing ToRs, based on the ones we have; and finally we flag out anomalous ToRs, some of which are clear suspects of complex ToRs. Some of the anomalies we find in the last stage are errors in our input datasets, which are then corrected for future use. Thus, as we keep using the analysis method periodically, we end up flagging only true anomalies and new changes in the Internet ToRs (like a new merger between two ASes). A detailed description of the method stages is as follows:

A. Preprocessing

The algorithm receives as inputs:

1) A dataset of IP-level traceroute measurements.
2) A mapping of IP addresses to PoPs and ASes.
3) A dataset providing initial classification of AS-level links ToR: c2p / p2c / p2p / s2s.

The first step of the preprocessing is identifying for each IP address in the traceroutes dataset its corresponding PoP and AS. IP addresses whose AS can not be identified (i.e., internal IP addresses) are discarded. IP addresses whose AS is known but their PoP is not are retained. The next step is identifying the path's AS borders, by finding pairs of consecutive hops which belong to different ASes. All IP-level hops which are not located on AS-AS border links are discarded, as we are only interested in links between different ASes. Finally, the algorithm discards repeating paths and paths which are fully contained within other paths. This last stage reduces the amount of paths, leaving only paths that contribute new information over others.

Traceroute paths may contain PoP loops or cycles, caused by load balancing artifacts, misconfigured routers or measurements taken during routing convergence periods [12]. For any path that contains a loop, the algorithm trims the path's prefix and suffix in order to retrieve the longest possible segment which does not contain a loop. We discard traceroute measurements which, when repeatedly measured, show artifacts of load-balancing routers.

The last step in the preprocessing stage is discarding IXP hops from the traceroutes. As some IXPs appear on traceroute paths as an additional AS hop, they may introduce errors in the following phases. Thus, if hop N in the traceroute represents an IXP, we drop this hop and stitch hops $N-1$ and $N+1$ together, forming as AS-to-AS level link. We further discuss the reason and effect of this step in section IV.

B. ToR augmentation

The ToR augmentation method, which is based on ideas from [7] and conducted on AS level, assumes validity of the valley-free rule on existing paths and infers new ToRs in a way which preserves this rule. This assumption is used to assign a ToR to links that have no ToR classification in the initial ToR database (See Section IV-A for the initial ToR database coverage).

To find the ToR of unclassified links, we consider AS-level link paths generated in the preprocessing stage. Only AS-level link paths that have a single unclassified link and that are otherwise valley-free are considered. For each undetermined link in a given path, a vote is cast for each type of ToR which will not violate the valley-free path property: A c2p vote is cast for links which are in the middle of an "uphill" segment or links between an uphill segment and a p2p link. A p2c vote is cast for links which are in the middle of an "downhill" segment or links between a p2p link and downhill segments. For links which are before downhill segments in a path where only a downhill segment is detected, or which are after uphill segments in a path where only an uphill segment is detected, or for links which are located exactly between the uphill and downhill segments, all three possible votes are casted: c2p, p2p and p2c.

After traversing all eligible paths, a new ToR is inferred for cross-AS PoP-links that had no ToR assigned. Such a link is assigned a ToR if the percentage of votes which agreed on a ToR is larger than a VOTING-THRESHOLD, and there were more than MIN-VOTES votes for the ToR. In case that multiple ToRs pass the above thresholds, we give precedence to the p2p ToR. The process is then repeated, taking newly discovered ToRs into consideration, and trying to infer ToR for the remaining unassigned links, until no new ToRs are discovered.

C. Complex ToRs and anomaly detection

A path which is not valley-free and can be corrected by changing a single link's ToR, is termed a *single-error path.* Single-error paths always contain one or two links whose ToR can be changed in order to make the path valley-free (a proof is omitted due to space limitation). These links are denoted *candidate anomalous links*. Each candidate anomalous link has one or two alternative ToRs: the ToRs which if assumed will make the path valley-free. For each PoP-PoP link A-B, the algorithm finds:

1) P: the group of paths that link A-B is part of.
2) n: the overall number of unique PoP and AS nodes in the graph created by combining all the paths that contain link A-B. A large number means A-B was measured by traceroutes with many diverse sources and destinations.
3) VP: the group of valley-free paths A-B is part of.
4) FP_{c2p}: the group of paths which are not valley-free, in which A-B is a candidate anomalous link, who can be

made valley-free by assuming c2p ToR for A-B. FP_{p2p} and FP_{p2c} are defined similarly.

The algorithm outputs anomalous PoP-PoP links which satisfy the following conditions:

- The link has a minimal measured graph size ($n > min\text{-}nodes$)
- The percentage of valley-free paths containing the link is smaller than an arbitrary *min-valid-percentage* ($|VP|/|P| < min\text{-}valid\text{-}percentage$)
- There is a new ToR which, when assumed for the link, turns a large percentage of paths to be valley free ($|FP_{ToR}|/|P| > min\text{-}fixed\text{-}percentage$)

The three conditions capture cases when a PoP-PoP link has a significant evidence for a problem (first two conditions) and a fix in the link ToR, which seems to correct the problem. The algorithm also outputs the set of PoP-level links which comply with the first two rules, but for which a new ToR could not be determined with a high level of confidence.

III. Datasets

Three types of datasets are used in this study:

DIMES traceroutes All the traceroutes measurements are taken from the DIMES project [13], from May 2012, weeks 19 and 20. The dataset includes 29.2 million traceroute measurements and 506.3 million IP-level hops. The measurements targeted 2.39 million destination IP addresses and were collected by 1017 DIMES agents. RouteViews [14] and WHOIS databases were used to infer every IP address to an AS.

DIMES PoPs The IP to PoP mapping dataset is taken from the DIMES project, from weeks 19 and 20 of 2012. The mapping was based on traceroutes taken by both DIMES and iPlane [10] over the same period of time. The map contains 5215 PoPs and 98650 IP addresses in 2636 different ASes. It is publicly available on the DIMES project website.

CAIDA ToRs The initial AS ToR mapping dataset is taken from CAIDA's AS Rank Website[1] from August 2012. The dataset relies on BGP paths obtained on June 2012. It contains ToRs for 119,924 AS couples. 76781 (64%) relationships are customer/provider relationships, 40,900 (34%) are peering relationships and 2243 (2%) are sibling relationships. We compare our results with a newer CAIDA dataset, published September 2012.

IV. Experimental Results

A. Preprocessing Results and ToR augmentation

The preprocessing stage of the algorithm takes the 29 million IP level traceroute measurements and turns them into 1.63 million unique PoP level paths, thus reducing the dataset size by an order of magnitude. 1.48 million PoP-level Paths (91%) are valley free.

Out of the 70714 AS-AS links found in the dataset, only 45202 links (64%) were covered by the CAIDA dataset. It is thus necessary to augment the ToR dataset. We complete the missing ToR for 6699 links and fail to complete 18813 AS-AS links, out of them 495 appear only on paths which are not valley free. Links of unknown ToR which appear only on paths which are not valley-free can not be assigned a ToR with a high level of confidence. The augmentation increases the number of customer-provider PoP links but only slightly increase the number of peer-peer links. For the ToR voting, we use a VOTING-THRESHOLD of 80%, which gives a high level of confidence that the inference is correct. We select this value based on experimentation with a range of values and find that the effect on the results is marginal. Further information is omitted due to space limitations.

The ToR augmentation method requires a minimal number of paths in which the inferred AS-AS link is included and that are valley-free, the MIN_VOTES threshold. This parameter is required as inferring ToRs according to a few paths might introduce errors due to wrong traceroute replies or wrong AS prefix resolution, similarly to the phenomena described by Zhang *et al.* [15]. We tested a range of MIN-VOTES values, in order to select the best threshold and to verify sensitivity. Setting MIN-VOTES$=5$ infers 6699 new ToRs, while MIN-VOTES$=3$ helps inferring 8594 new ToRs. However, for a large number of AS-AS links there is only a single applicable path (regardless of the valley free rule), which makes the augmentation difficult. Under such conditions we do not attempt to infer the ToR. For lack of space we omit further discussion of MIN-VOTES sensitivity. We eventually set MIN-VOTES$=5$ which is a high confidence threshold, and allows to infer 26% of the missing ToRs.

B. Sensitivity analysis

Two parameters affect the anomaly detection method. The first, *min-valid-percentage*, determines the minimal percentage of valley-free paths required to consider the PoP-PoP ToR correct (as detailed in Section II-C). The second parameter, *min-fixed-percentage*, determines the minimal percentage of valley-free paths after the ToR was replaced required to consider the new PoP-PoP ToR correct. We evaluate the effect these two parameters have on our anomaly detection method's results.

min-valid-percentage and *min-fixed-percentage* capture the amount of confidence we wish to achieve in determining whether a specific PoP-PoP link is anomalous. A larger *min-valid-percentage* may cause non-anomalous links to appear as anomalous, but can also lead to the discovery of anomalous links that by chance did not consistently cause path invalidity. A low *min-fixed-percentage* threshold marks PoP-PoP links that even after changing their ToR appear as candidates to be anomalous due to non valley-free paths. This may happen when some of the paths contain other errors, such as traceroute measurement errors resulting from wrong AS resolution or ToR errors on other AS-AS links.

To study the sensitivity to thresholds, we omit anomalies that turn out to be errors in the original AS ToR database or that are caused by IXPs. This is done as these are one-time corrections and do not affect the algorithm in later runs. Figure 1 shows the effect of changing the two parameters on

[1] $http://www.caida.org/data/active/asrelationships/$

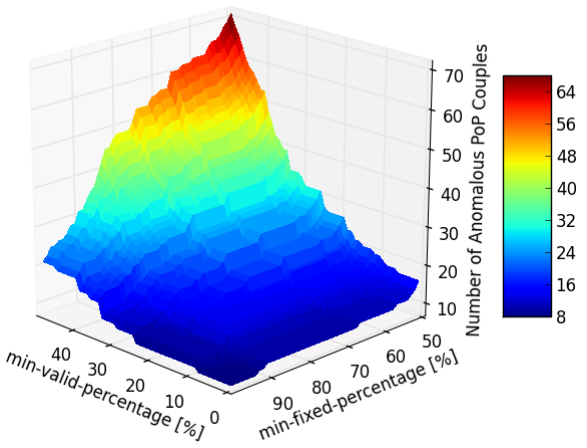

Figure 1. Anomaly Detection vs. Thresholds Values.

the number of discovered anomalous PoPs, with *min-nodes* set to 10 nodes. For the purpose of sensitivity study, we consider as anomalous PoPs only PoPs that fall under the categories of complex AS relationships and odd academic ToRs (see below). Clearly for a large range, between 0% and 35% for the *min-valid-percentage* threshold and between 70% and 100% of the *min-fixed-percentage* threshold, there is little change in the number of discovered anomalies. Thus, We select the thresholds from the non-sensitive region: *min-valid-percentage*$= 20\%$ and *min-fixed-percentage*$= 75\%$.

An interesting observation is that eight PoP couples are "perfect anomalies": they appear in no valid PoP paths, but when changing their ToR all the paths in which they appear become valley-free.

C. Anomaly detection

After the first execution of the anomaly detection algorithm, we detect a couple of dozens anomalies. We classify these anomalies into seven categories and highlight specific cases that exemplify the anomaly type:

1) AS Prefix Resolution Errors: Our anomaly detection method detected three cases that were attributed to AS prefix resolution errors. In these cases, the corresponding AS for a specific IP address in a traceroute measurement was incorrectly resolved by the RouteViews dataset. This caused a large percentage of the paths which contained this address to contain a valley, as the ToR between the wrongly assigned AS and its neighbors was incorrect. AS prefix resolution errors might occur when the BGP blocks that were announced to RouteViews were incorrect or not updated. Closer inspection, using other tools including WHOIS, revealed the true owner of the IP address. Assessing the accuracy of multiple IP to AS resolution databases is outside the scope of this paper.

Figure 2 demonstrates this phenomenon. In this case, an anomalous link is detected between AS2116 (Ventelo) and AS3549 (Global Crossing), AS3549 is the provider according to CAIDA. In all the paths that contained this link, it appeared after a link between AS3356 (Level 3) and AS2116 (Ventelo), which is a p2c link, creating a valley in these paths. However, using WHOIS it was discovered that the IP prefix to AS

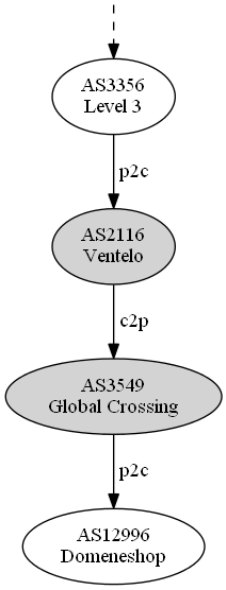

Figure 2. AS Prefix Resolution Error - Example

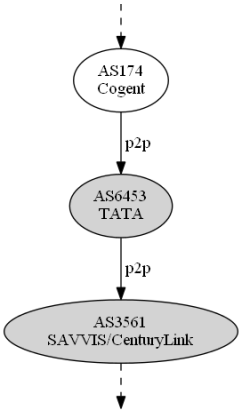

Figure 3. Complex AS Relationship - Example

mapping was wrong, and that the PoP first associated with AS3549 actually belongs to Domeneshop (AS12996), which is a customer of Ventelo.

2) IXP and sibling detection: Usually, when IXPs appear in traceroute paths it is as an additional IP hop. In ToR analysis they should be removed or else introduce errors since they are not part of the AS hierarchy, which we did in our preprocessing stage using lists of known IXPs. However, we have found six IXPs that appeared as anomalies in our PoP level traceroutes. Finding IXPs and consequently other anomalies is an incremental process, as each detected IXP allows more paths to become valley-free (due to their omission).

Similarly, we detected wrongly inferred siblings relationships. These are often cases of one ISP taking over a second ISP, which was previously its customer. This change of ToR is not always updated in the ToR dataset. Thus, when checking valley free routing, some of the paths between the pair of ASes will remain valid as c2p, while others will only be valid as s2s. Since in many routes a s2s ToR is interchangeable with c2p ToR, the change of ToR between the two ASes may be hard to detect. We manage to find 6 wrongly inferred s2s relationships, e.g., between TelePacific (AS14265) which acquired Mpower (AS18687).

3) ToR inference errors: On three cases, a PoP-PoP link was deemed anomalous, but closer inspection revealed that the ToR for the corresponding AS-AS link was wrongly inferred by CAIDA. In general, the method tries to avoid flagging such cases as anomalies. It does so by discarding anomalous candidate links for which the confidence for corresponding ASes' ToR is not high enough. We deem two ASes' ToR as confident if there is a majority of paths containing the AS-AS link which follow the valley-free rule.

Two of the three cases we've discovered were corrected in CAIDA's September 2012 ToR dataset, a few weeks following

this analysis. In the third case, CAIDA inferred a peering relationship between two ASes (AS12389 and AS8359), while in our measurements almost half of the paths which contained this AS-AS link were not valley free.

One exemplary case of a wrongly inferred ToR, CAIDA inferred the AS3561-AS4134 (SAVVIS-Chinanet) as a peering relationship. Our algorithm detected specific PoPs belonging to these organizations as anomalous, and suggested a p2c relationship instead. The PoPs were deemed anomalous as there was a small majority of paths containing PoP couples from AS3561 and AS4134 (80 out of 145 paths) which were valley free. In September 2012 CAIDA updated this ToR in their dataset and changed the relationship between the two ASes to p2c, same as suggested by our algorithm.

4) Complex AS relationships: An interesting relationship was found between two PoPs of AS3561 (SAVVIS/CenturyLink) and AS6453 (TATA) in Canada. As both ASes are Tier-1 providers, the assumed ToR between them is a peering relationships (also indicated by CAIDA). However, only three out of the sixteen unique PoP paths that include a link between this pair of PoPs are valley-free. Out of the remaining 13 paths, 11 traverse a PoP link between AS174 (Cogent) and AS6453 (TATA), clearly another p2p link between tier-1 ASes (see Figure 3). When assuming a c2p relationship (the provider being AS6453's PoP), all paths are valley-free.

It seems that while CenturyLink and TATA have a peering relationship in most locations, this specific PoP-PoP link is configured differently: TATA's specific PoP provides transit services between other Autonomous Systems (namely, Cogent) and SAVVIS.

We have revalidated this finding a couple of weeks after the original experiment's date, by running a dedicated DIMES experiment that, issuing a large amount of traceroute measurements towards the specific IP addresses of these PoPs from many widely spread vantage points. The phenomenon was also reproducible by issuing traceroutes from Cogent routers, using their Looking Glass service.

5) Odd Academic ToRs: A couple of ToR anomalies are discovered in research institutes' affiliated PoP links. Research institutes are less driven by commercial incentives and tend to be more collaborative in nature, thus setting their ToR criteria differently than most ASes.

Figure 4 shows one such case, involving multiple PoPs belonging to research organizations. Traffic flowed from multiple PoPs belonging to SWITCH, the Swiss Education and Research Network (AS559) through CERN (AS513) and then through KPN (AS286), finally reaching the tier-1 provider Level3 (AS3356). According to the ToRs inferred by CAIDA, SWITCH is a provider of CERN, and KPN is a peer of Level3, causing this path to be non valley-free. CAIDA's dataset missed information on the ToR of CERN and KPN, but for any ToR this path violates the valley-free rules.

An additional anomaly, shown in Figure 5, is a single HP (AS71) PoP that is connected to the Internet via Stanford University (AS32) and CSUNET (AS2153). CAIDA's ToR

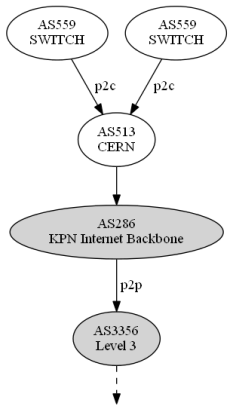

Figure 4. Academic ToR Anomaly - First Example

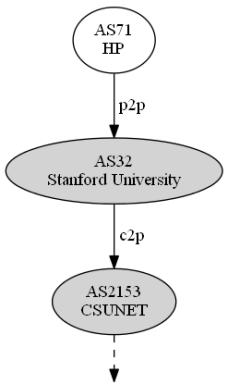

Figure 5. Academic ToR Anomaly - Second Example

for the HP-Stanford link was p2p, and the Stanford-CSUNET link was c2p (Stanford is the customer), resulting in a clear anomalous link.

6) Traceroute errors: We have found three cases in which we believe detected anomalies were probably caused by wrong router replies. In these cases, a specific IP address which was part of a reported traceroute path was not the actual IP address traversed by traffic on this path. This is caused by ICMP replies that are sent through a different interface than the one the packets actually went through [16]. This error may result in a wrong AS resolution, leading to wrongly assumed non-valley free paths.

7) Unresolved Anomalies: Three of the anomalies we found remain unresolved. While we have assumptions for the nature of these anomalies, we did not manage to corroborate our finding and thus prefer to declare them unclassified.

D. Discussion

1) ToR Datasets: AS ToR datasets fail to capture a large amount of AS links, due to their reliance on specific data sources. For example, CAIDA's AS Relationships Dataset [17] only uses BGP routes in order to infer AS ToRs. Shavitt and Shir [13], and more recently Gregori *et al.* [18], showed that BGP and traceroute measurement sources complement each other. Therefore, our augmentation of an existing AS ToR dataset according to an additional source of traceroute measurements is important by itself.

2) ToR inference at different layers of aggregation: To better understand the contribution of PoP level maps to ToR research, the disadvantages in using other levels of aggregation should be discussed in comparison.

For many years, using the AS level graph to infer ToRs seemed to be the right way, as this is seemed to be the level at which ToRs are defined. In addition, the AS graph is relatively small and easy to study. However, the AS level treatment does not allow the inference of complex relationships, where two ASes have different relationships in two different locations. As demonstrated by Dimitropoulos *et al.* [6], ASes might have a more complex relationship in various peering points. In addition, most existing algorithms use specialized methods

for sibling relationship detection, which rely on data sources other than BGP and traceroute measurements [17].

Using router or IP level maps for complex relationship detection is also not a good solution as it is hard to identify in them scattered errors and anomalies. IP level paths introduce noise to the measurements and cause anomalies to be dispersed over multiple IP addresses, diminishing their significance and preventing their accurate detection. In addition, router and IP level datasets are very large and require considerable processing resources.

PoP level maps provide an answer to the above issues and propose a better level of aggregation than AS, Router or IP level for anomaly detection. If two ASes have different relationships in two different locations, these will be represented by two distinct PoP-PoP links, and one of them will clearly violate the valley free rules, and thus can be easily flagged. While the same information will also be detectable on the IP/router level, it will be hard to correlate it to a specific location and to discard local errors. Considering the same problem the other way around, when detecting on the IP/router level multiple non valley free routes it is hard to understand the nature of each link's anomaly or error. The aggregation of multiple IP/router level links to a single PoP-PoP link reduces the complexity of this issue considerably, and provides a higher level of confidence to the inferred new ToR.

3) Dataset Size Dependence: ToR Errors and PoP-PoP link anomalies are more likely to be found as we increase the number of measurements and diversify the measurement vantage points. When measured from a small number of sources, an anomaly might not be identified, since a single path might not violate valley-freedom even with the existing error or anomaly. For example, if a ToR is inferred as p2p instead of c2p, it might not be discovered if the link resides between a c2p link and a p2c link. These paths would be valley-free in both cases, and our method - which looks for improvement in the percentage of valley-free paths when assuming a different ToR - would not identify this anomaly.

It is important to note that anomalies detected in a given dataset on the PoP level will remain valid even if the dataset grows considerably, since the threshold to flag an anomalous ToR is based on the number of violating PoP level paths and not their percentage.

4) Validation: The validation of our method is a hard task. Except for verifying results with ISPs, which are reluctant to cooperate, there is no single ground truth dataset. As we show, many of the datasets that we use as a reference have errors. Some of our results are corroborated by corrections done to the CAIDA dataset shortly after we ran our analysis. Another mean of validation is from ISPs websites and public information. This applies mainly for siblings ToR validation, often caused by one ISP acquiring another.

Another method of validation is using targeted measurements through many scattered vantage points to the point of anomaly. This is intended to eliminate transient routing effects and to confirm the anomaly through as many distinct paths as possible. For some anomalies, such as mistakes in AS resolution, reverse DNS and WHOIS, are useful tool in finding the true IP to AS mapping.

We believe that the level of validation provided in this work is sufficient under the given lack of ground truth conditions and as the results show, it provides a good mean to validate other datasets and sources for ToR information.

V. Conclusions

In this paper we presented a method to infer AS relationships using PoP data. The method is useful to detect complex types of relationships as well as anomalies and mistakes in existing ToR datasets. The method leverages PoP-level maps, which reduces the size of the analyzed datasets and highlights anomalies that are otherwise hard to detect on the IP or the router level. In the future, we intend to extend this study, further examining and validating complex AS relationships and anomalies. Additional future work will focus on geography related aspects of ToR, and how they affect the robustness of the network.

References

[1] G. Huston, "Interconnection, peering, and settlements," in *INET*, San Jose, CA, USA, Jun. 1999.

[2] L. Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM Transactions on Networking*, vol. 9, no. 6, 2001.

[3] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the internet hierarchy from multiple vantage points," in *IEEE INFOCOM*, New York, NY, USA, Apr. 2002.

[4] G. D. Battista, M. Patrignani, and M. Pizzonia, "Computing the types of the relationships between autonomous systems," Dipartimento di Informatica e Automazione, Universita di Roma Tre, 2002., Tech. Rep. RT-DIA-73-2002, 2002.

[5] T. Erlebach, A. Hall, and T. Schank, "Classifying customer-provider relationships in the internet," in *CCN*, 2002.

[6] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, kc claffy, and G. Riley, "AS relationships: Inference and validation," *ACM SIGCOMM Computer Communications Review*, vol. 37, 2006.

[7] Y. Shavitt, E. Shir, and U. Weinsberg, "Near-deterministic inference of AS relationships," in *ConTEL 2009*, Jun. 2009, pp. 191 –198.

[8] D. G. Andersen, N. Feamster, S. Bauer, and H. Balakrishnan, "Topology inference from bgp routing dynamics," in *IMW '02*, 2002, pp. 243–248.

[9] N. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP topologies with Rocketfuel," in *SIGCOMM '02*, 2002, pp. 133–145.

[10] H. V. Madhyastha, T. Anderson, A. Krishnamurthy, N. Spring, and A. Venkataramani, "A structural approach to latency prediction," in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, ser. IMC '06, 2006, pp. 99–104.

[11] D. Feldman, Y. Shavitt, and N. Zilberman, "A structural approach for PoP geolocation," *Computer Networks*, vol. 56, no. 3, 2012.

[12] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with paris traceroute," in *IMC'06*, 2006, pp. 153–158.

[13] Y. Shavitt and E. Shir, "DIMES: Let the Internet measure itself." in *ACM SIGCOMM Computer Communication Review*, vol. 35, Oct. 2005.

[14] "University of Oregon RouteViews Project." [Online]. Available: http://www.routeviews.org/

[15] Y. Zhang, R. Oliveira, H. Zhang, and L. Zhang, "Quantifying the pitfalls of traceroute in as connectivity inference," in *Passive and Active Measurement*, ser. Lecture Notes in Computer Science, 2010, vol. 6032.

[16] Y. Hyun, A. Broido, and k. Claffy, "On Third-party Addresses in Traceroute Paths," in *PAM'03*, San Diego, CA, 2003.

[17] "The CAIDA AS Relationships Dataset." [Online]. Available: http://www.caida.org/data/active/as-relationships/

[18] E. Gregori, A. Improta, L. Lenzini, and C. Orsini, "The impact of ixps on the as-level topology structure of the internet," *Computer Communications*, vol. 34, no. 1, pp. 68 – 82, 2011.