Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers

Bradley Huffaker, Amogh Dhamdhere, Marina Fomenkov, kc claffy
{bradley,amogh,marina,kc}@caida.org *

CAIDA, University of California, San Diego

Abstract. To describe, analyze, and model the topological and structural characteristics of the Internet, researchers use Internet maps constructed at the router or autonomous system (AS) level. Although progress has been made on each front individually, a *dual graph* representing connectivity of routers with AS labels remains an elusive goal. We take steps toward merging the router-level and AS-level views of the Internet. We start from a collection of traces, i.e. sequences of IP addresses obtained with large-scale traceroute measurements from a distributed set of vantage points. We use state-of-the-art alias resolution techniques to identify interfaces belonging to the same router. We develop novel heuristics to assign routers to ASes, producing an **AS-router dual graph**. We validate our router assignment heuristics using data provided by tier-1 and tier-2 ISPs and five research networks, and show that we successfully assign 80% of routers with interfaces from multiple ASes to the correct AS. When we include routers with interfaces from a single AS, the accuracy drops to 71%, due to the 24% of total inferred routers for which our measurement or alias resolution fails to find an interface belonging to the correct AS. We use our dual graph construct to estimate economic properties of the AS-router dual graph, such as the number of internal and border routers owned by different types of ASes. We also demonstrate how our techniques can improve IP-AS mapping, including resolving up to 62% of false loops we observed in AS paths derived from traceroutes.

1 Introduction

There is growing scientific interest in the structure and dynamics of Internet topology, primarily at the router and Autonomous System (AS) levels. Substantial progress over the last decade toward understanding and improving the integrity and completeness of router and AS-level topologies *separately* (reviewed in Section 4) has inspired us to seek a graph construction that *merges router and AS-level views of the Internet*. Such a view would capture administrative boundaries while providing sufficient detail about the geography and internal structure of each AS. Inherent limitations and inaccuracies of existing techniques for alias resolution, IP-to-AS mapping, and router-to-AS assignment (not to mention validation of any of them) render this goal challenging.

In this work we take initial steps toward merging router and AS-level views into a *dual graph* representation of the Internet. We start from active measurement (traceroute-like) datasets collected using CAIDA's Archipelago distributed measurement infrastructure (Ark) [17]. We then apply state-of-the art alias resolution techniques [19] to infer

* Support for this work is provided by DHS N66001-08-C-2029 and NSF 05-51542.

which interfaces belong to the same router, creating a router-level Internet map. Finally, we propose heuristics to assign routers to ASes, using information derived from the interfaces that we infer belong to a particular router. We evaluate our AS assignment heuristics by validating against ground truth data from tier-1 and tier-2 ISPs and five research networks. We successfully assigned 80% of multi-AS routers, i.e, routers whose interfaces map to different ASes. When we include single-AS routers (routers whose interfaces all map to the same AS), the accuracy drops to drops to 71%, due to the 24% of total inferred routers for which our measurement or alias resolution fails to find an interface belonging to the correct AS. We also demonstrate how our techniques can be used to study the statistical properties of the resulting AS-router dual graph, and can improve IP-AS mapping of state-of-the-art AS-level traceroute tools.

2 Datasets and methodology

We briefly describe three components of our methodology: gathering a large set of Internet path data; resolving IP address aliases to create a router-level graph; and designing heuristics to map annotated routers to ASes. All CAIDA data sets and tools developed to support this work will be publicly available.

2.1 Datasets

Active measurements
We collected our active measurements using CAIDA's Archipelago (Ark) Measurement infrastructure [17], using 37 monitors in 28 countries. The Ark monitors used Paris traceroute [6] to randomly probe destinations from each routed /24 seen in BGP dumps from Routeviews over a 28-day collection period in September and October 2009. We call the resulting set of 268 million traceroute paths our *traceroute* dataset, which we used to infer which IP interfaces belong to the same router (Section 2.2).

BGP data
To assign IP addresses to ASes, we used publicly available BGP dumps provided by Routeviews [26] and one of RIPE NCC's collectors (RCC12) [25]. BGP (Border Gateway Protocol) is the protocol for exchanging interdomain routing information among ASes in the Internet. A single origin AS typically announces ("originates") each routable prefix via BGP. We perform IP-to-AS mapping by assigning an IP address to the origin AS of the longest matching prefix for that IP address. We also used this BGP data to annotate each interdomain link with one of three (over-simplified) business relationships: customer-provider (the customer pays the provider); settlement-free peer (typically no money is exchanged); and sibling (both ASes belong to the same organization) – using the classification algorithm in Dimitropolous *et al.* [10].

Ground truth dataset
Our ground truth datasets includes private data from a tier-1 ISP (ISP_1) and a tier-2 ISP (ISP_2). In addition we use public data from the following research networks: CANET (ISP_C)[1], GEANT (ISP_G)[2], Internet2 (ISP_T)[4], I-Light (ISP_L)[3], and National LambdaRail (ISP_N)[5]. ISP_1 and the five research networks provided the full list of interfaces. ISP_1 and ISP_2 provided their hostname conventions, which allowed us to

identify interfaces in their address space, but not on their routers. We thus have two sets of interfaces for each network i: $\mathcal{I}_i$ (interfaces on routers that belong to network i) and $\bar{\mathcal{I}}_i$ (interfaces in i's address space, but on routers that do not belong to network i). For each network we then generate a list of AS numbers known to belong to that network: $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_C$, $\mathcal{A}_I$, etc., and the set of ASes that are not in each $\mathcal{A}_i$, denoted $\bar{\mathcal{A}}_i$.

2.2 Alias resolution

For alias resolution, we rely on CAIDA's alias resolution tools MIDAR and *kapar* [19]. MIDAR expands on the IP velocity techniques of RadarGun [8], and kapar expands on the analytical techniques of APAR [14]. We first use the *traceroute* dataset as input to MIDAR, the output of which is fed into *kapar*. *kapar* heuristically infers the set of interfaces that belong to the same router, and the set of two or more routers on the same "IP link" (which could either be a point-to-point link, or LAN or cloud with multiple attached IP addresses). *kapar* produces two datasets corresponding to inferred nodes (routers) and links. Each node in the router dataset has a set of *known interfaces* and *inferred interfaces*. Known interfaces were measured directly; inferred interfaces result from *kapar* determining that a router r_1 has a link to interface i_2 on router r_2, but we did not see an actual interface on router r_1. The interfaces on an IP link are typically assigned IP addresses from the same prefix, so we assume that router r_1 must have an interface *from the same prefix as* i_2. The link dataset contains, for each link, the set of routers and router interfaces that we inferred as sharing that link. *kapar* correctly identified 66% of the true aliases from among the set of ISP_1's observed interfaces (our largest set of ground truth data), with a 5% false positive rate.

At least three limitations of our alias resolution techniques may affect the AS assignment process. First, a large number of interfaces and links between them are never observed, either because they do not respond to ICMP, or because none of the traceroutes encounter those interfaces. Second, some interfaces that respond to ICMP have addresses belonging to private address space, which makes them indistinguishable from other interfaces using the same private address space. Third, even when all of a router's interfaces are discovered, we may have insufficient information to infer that they belong to the same router. For example, we inferred 1390 routers as having interfaces from a single AS in $\mathcal{A}_1$, which our method would infer to mean these routers are in ISP_1. But our ground truth dataset refutes this inference; these routers do not belong to ISP_1, and likely have an interface (which we either did not observe or did not resolve accurately) from at least one other AS in $\bar{\mathcal{A}}_1$.

2.3 AS assignment Methods

The goal of the AS assignment process is to determine the AS that owns each router. For each router r, we create an *AS frequency matrix* that counts the number of interfaces (known and inferred) from each AS that appears on r. The ASes in this frequency matrix represent the set of possible owner ASes of r. Next, we describe the heuristics we designed to determine r's ownership from among the candidates present in r's AS frequency matrix. Figure 1 illustrates the five heuristics examined in this paper.

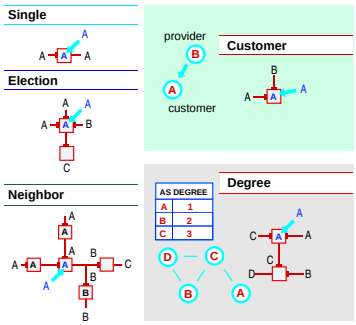

Fig. 1: Depiction of five evaluated heuristics for assigning AS labels to routers: **Single** (only one choice); **Election** (assign to AS with largest number of interfaces); **Neighbor** (assign to AS with most neighbors); **Customer** (assign to customer AS); **Degree** (assign to smallest degree AS).

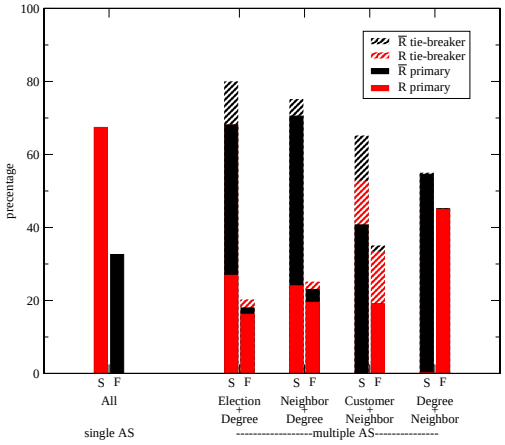

Fig. 2: Success (S) and failure (F) rates of AS assignment primary heuristics, and the best tie-breaking heuristics for each primary, for single-AS and multi-AS routers in $\mathbb{R}$ and $\bar{\mathbb{R}}$.

Single: This heuristic is used for the case where a single AS is present in r's AS frequency matrix. In this case, we (trivially) assign r to this AS.

Election: This heuristic assigns a router r to the AS with the highest frequency in r's AS frequency matrix, assuming routers tend to have more interfaces in the address space of their owner. **Election** produces an ambiguous assignment when multiple ASes have the same (highest) frequency, which occurred for 14% of the multi-AS routers in our set.

Neighbor: For this heuristic, we first determine the set of single-AS routers to which r is connected (its single-AS neighbors). We create a new AS frequency matrix that counts the number of single-AS neighbors of r from each AS. The **Neighbor** heuristic assigns r to the AS with the largest frequency (most single-AS neighbors), based on the intuition that a router is connected to a larger number of single-AS routers in its owner AS. **Neighbor** produces an ambiguous assignment when multiple ASes have the same (highest) frequency.

Customer: This heuristic uses the AS relationship dataset to assign relationships to each pair of ASes from r's AS frequency matrix[1]. **Customer** assigns r to the AS inferred to be a customer of every other AS in r's AS frequency matrix. This heuristic is based on the common practice that customer and provider routers typically interconnect using addresses from the provider's address space. Consequently, a router with interfaces from both the customer and provider address spaces is assigned to the customer.

[1] Not every possible AS pair in r's frequency matrix has a known relationship; many AS pairs have no link between them in the original BGP AS graph, so no defined relationship.

Degree: For this heuristic, we first generate an AS-level graph by assuming full-mesh connectivity among ASes from each router's AS frequency matrix. We then use this graph to generate an AS degree for each AS. **Degree** assigns router r to the smallest-degree AS from r's AS frequency matrix, i.e., the AS most likely to be the customer AS, based on similar intuition as the **Customer** heuristic.

2.4 Evaluation of AS Assignment heuristics

We next evaluate our AS assignment heuristics by comparing our AS assignment with our ground truth datasets. We classify each router inferred by *kapar* into the following sets. If a router r_0 has at least one interface in $\mathcal{I}_i$, then we assign r_0 to the set $\mathcal{R}_i$ (the set of routers owned by ISP$_i$). If a router r_1 has at least one interface from the set $\bar{\mathcal{I}}_i$, then we assign r_1 to the set $\bar{\mathcal{R}}_i$ (inferred routers not owned by ISP$_i$). We found 39 routers (0.6% of the total analyzed) with interfaces in both $\mathcal{I}_i$ and $\bar{\mathcal{I}}_i$ or $\mathcal{I}_i$ and $\mathcal{I}_j$, which contradicts the meaning of these data sets (describing mutually exclusive routers). These discrepancies are due to false positives in our alias resolution process, so we discard them for the purpose of evaluating our AS assignment heuristics. All but three routers in $\mathcal{R}_i$ have a single AS in $\mathcal{A}_i$ ($\mathcal{A}_i$ is the set of ASes owned by ISP$_i$), which means there is a single successful assignment for most routers. For the three routers with multiple ASes in A_i, successful assignment is ambiguous, and we omitted these routers from the evaluation, leaving us with $|\mathcal{R}_1|$ = 3,405 and $|\bar{\mathcal{R}}_1|$ = 2,254, $|\mathcal{R}_2|$ = 241 and $|\bar{\mathcal{R}}_2|$ = 86, $|\mathcal{R}_G|$ = 37 and $|\bar{\mathcal{R}}_G|$ = 0, $|\mathcal{R}_L|$ = 32 and $|\bar{\mathcal{R}}_L|$ = 0, $|\mathcal{R}_T|$ = 17 and $|\bar{\mathcal{R}}_T|$ = 0, $|\mathcal{R}_N|$ = 16 and $|\bar{\mathcal{R}}_N|$ = 0, and $|\mathcal{R}_C|$ = 8 and $|\bar{\mathcal{R}}_C|$ = 0. We call the combined set of all routers $\mathbb{R} = \cup\mathcal{R}_i$, those owned by some network in our ground truth dataset, and the set $\bar{\mathbb{R}} = \cup\bar{\mathcal{R}}_i$ those we know not to be owned by a specific network in our ground truth datasets. Using our knowledge of interface ownership, we derive $|\mathbb{R}|$ = 3,795 and $|\bar{\mathbb{R}}|$ = 2,340 routers on which to test AS assignment heuristics. We consider $H(r)$, the AS to which a certain heuristic assigns router r, as a *successful assignment* if $((r \in \mathcal{R}_i)\&\&(H(r) \in \mathcal{A}_i))||((r \in \bar{\mathcal{R}}_i)\&\&(H(r) \in \bar{\mathcal{A}}_i))$, i.e., if the router is in $\mathbb{R}$ and $H(r)$ selects an AS owned by the same ISP as the router, or the router is in $\bar{\mathbb{R}}$ and $H(r)$ selects an AS not owned by the ISP known to not own router.

Section 2.3 outlined the cases for which each heuristic provides an ambiguous assignment. To resolve ambiguous assignments, i.e., break ties, we paired each heuristic with a second one. We tested all combinations of pairs of heuristics to find the best tie-breaker[2] for each primary heuristic, resulting in the following combinations: **Election** + **Degree**, **Neighbor** + **Degree**, **Customer** + **Neighbor**, and **Degree** + **Neighbor**.

Figure 2 shows the fraction of routers we assigned successfully (bars labeled "S"), and the fraction that were failures (bars labeled "F"), determined using the ground truth datasets. Figure 2 presents these results separately for routers in $\mathbb{R}$ and $\bar{\mathbb{R}}$, and for different assignment heuristics. We found that for single-AS routers, all heuristics are either successful for the 67% in $\mathbb{R}$ or failures for the 33% in $\bar{\mathbb{R}}$. The explanation is straightforward: All routers in $\mathcal{R}_i$ or $\bar{\mathcal{R}}_i$ have at least one interface in ISP$_i$'s address space (not necessarily being used by ISP$_i$), and by extension an AS in $\mathcal{A}_i$. For single-AS routers in

[2] The best tie-breaker is the heuristic that produced the largest number of successful assignments for routers where the primary heuristic resulted in an ambiguous assignment.

$\mathcal{R}_i$, the AS must belong to $\mathcal{A}_i$, and the assignment is a success. For single-AS routers in $\bar{\mathcal{R}}_i$, assigning it to that single AS results in failure. For these single-AS routers in $\bar{\mathcal{R}}_i$, we have most likely failed to either see or accurately resolve the alias for the router's interface in address space not owned by ISP_i.

Figure 2 shows that when a router has interfaces from multiple ASes, the most effective stand-alone heuristic was **Neighbor**, which successfully assigned 70% of these routers. **Election** + **Degree** was the most successful combination of heuristics (mainly due to fewer failures on routers from $\bar{\mathbb{R}}$), with a success rate of 80%.

3 Applications of AS Assignment

In this section, we use the AS assignment heuristics described in Section 2.3 to produce a *dual graph* that merges router and AS-level topologies. We then describe two applications of this dual graph construct – producing representative dual topologies of the Internet, and improving the accuracy of AS-level traceroute tools.

3.1 Toward representative dual topologies of the Internet

Previous work [11, 21] has focused on generating AS-level graphs of arbitrary size, while preserving the correlation structure seen in real Internet topology, e.g., correlations between the number of customers, providers and peers of an AS, or between degrees of ASes at each end of an interdomain link. We seek to extend this previous work by designing a graph generator that can produce Internet-like dual topologies, i.e., AS annotated router-level graphs, of arbitrary size, preserving the statistical properties of the Internet's dual graph. Another application is to security-related situational awareness objectives, which require knowledge of the internal structure of ASes. We focus on two questions: How many inferred single-AS (internal) and multi-AS (border) routers do ASes own (with the aforementioned caveat that we may mis-characterize routers as single-AS if we undersample or mis-resolve interfaces)? Is there a correlation between an AS's degree and the number of routers it owns? We use the heuristics from Section 2.3 to assign routers to ASes, and measure the router ownership properties of resulting ASes. Our results do not represent the actual number of routers owned by an AS, only the number observed in our data samples.

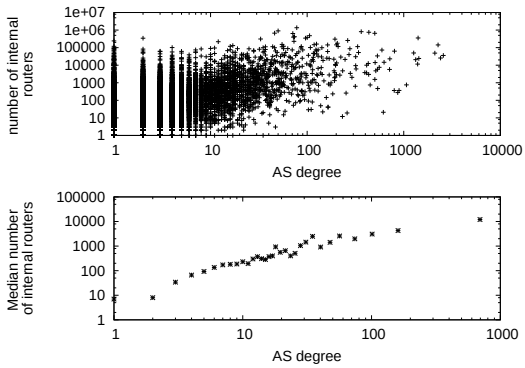

Fig. 3: The number of single-AS routers per AS vs degree (top) and the median number of single-AS routers per AS vs degree (bottom)

We first examine the number of single-AS routers owned by an AS, which does not depend on the assignment heuristic we used, since every heuristic assigns a single-AS

router to the same AS. The top graph in Figure 3 shows a scatter plot of the number of single-AS routers inferred per AS versus the AS degree as observed in BGP data (from Routeviews2 and RIPE's RRC12). We confirmed the expected positive correlation, where ASes with larger degrees (which typically represent larger transit providers) tend to have more single-AS routers. Several outliers have many single-AS routers and relatively low AS degrees (1 or 2). The top 10 such outliers corresponded to ASes that were either regional networks of a larger transit provider, or smaller administrative domains within a large transit provider. Consequently, these ASes had just one or two observed AS links, to the backbone AS of the larger transit provider. It is plausible that such regional transit networks or access provider networks have a large number of single-AS routers.

The bottom graph in Figure 3 shows the median number of single-AS routers per AS as a function of the AS degree. We bin ASes according to their degree, ensuring a minimum bin size of 50 ASes. We see a strong positive correlation between the number of single-AS routers and the inferred AS degree, which is expected since ASes with larger AS degrees typically represent transit providers, which need many routers. ASes with lower degrees are typically stub networks with less internal routing infrastructure.

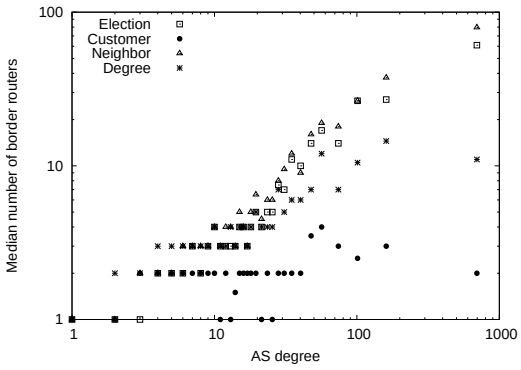

Fig. 4: Number of multi-AS routers per AS using **Election**, **Neighbor** and **Degree** heuristics shows strong correlation with AS degree.

Figure 4 shows the number of multi-AS routers owned by an AS as a function of AS degree, for different AS assignment heuristics. We found similar results with the **Election**, **Neighbor** and **Degree** assignment heuristics, and a strong positive correlation between the number of multi-AS routers of an AS and its degree. The **Customer** heuristic shows a much weaker correlation between the number of multi-AS routers and AS degree. **Customer** favors lower (BGP) degree ASes, since customer ASes tend to have smaller degrees than their providers, and **Customer** assigns a multi-AS router to the customer AS, decreasing the number of multi-AS routers for ASes with larger BGP degrees. We found that the **Neighbor** heuristic favors higher (BGP) degree ASes, inflating the number of multi-AS routers for higher degree ASes.

3.2 Toward accurate AS-traceroute

As a second application of the dual graph construct, we outline an approach to designing a more accurate AS-traceroute tool, a problem first studied by Mao *et al.* [23]. Mao *et al.* concluded that an accurate router-level map of the Internet would help to resolve anomalies seen in AS paths derived from traceroutes. Here, we investigate whether our AS assignment heuristics can improve AS-traceroute accuracy, by resolving anomalies such as missing AS hops, extra AS hops and AS loops. Identifying missing and extra

AS hops requires BGP feeds from the vantage points used for traceroute measurements, which the Ark infrastructure does not yet have. However, we can identify traceroutes that have AS loops, by performing an IP-to-AS mapping using BGP dumps collected from Routeviews and RIPE. Mao *et al.*. [23] noted two possible explanations for false loops in traceroute paths: the presence of Internet Exchange Point (IXP) infrastructure, and sibling ASes. We investigated whether our router-to-AS assignment alone can help to resolve these loops. In future work, we plan to incorporate IXP data collected by Augustin *et al.*. [7] to identify false loops due to IXP infrastructure, and WHOIS data to identify false loops due to sibling ASes.

By applying IP-to-AS mapping on the sequence of interfaces seen in each traceroute, we found that most Ark monitors yielded fewer than 5% of inferred AS paths that had loops. However, traces from one particular monitor yielded 75% of inferred AS paths with loops, which we discovered was caused by a single incorrectly mapped interface traversed by most traces from that monitor. We removed these traces for the remainder of our analysis. We then assigned an AS to each inferred router on the path using the AS assignment heuristics from Section 2.3. We replaced the loop segment in the traceroute AS path with an AS path segment derived from the router assignment heuristic. We measured the fraction of paths with traceroute loops resolved, i.e., removed, via this procedure. Figure 5 shows the fraction of traces with AS path loops that we could resolve using each of the AS assignment heuristics. We found that the **Customer** heuristic performed poorly. The **Neighbor** heuristic, which was the most accurate stand-alone AS assignment heuristic (Section 2.4) was able to resolve 62% of AS path loops. The combination **Election**+**Degree**, which was the most accurate combination AS-assignment heuristic, was able to resolve just over 61% of AS path loops.

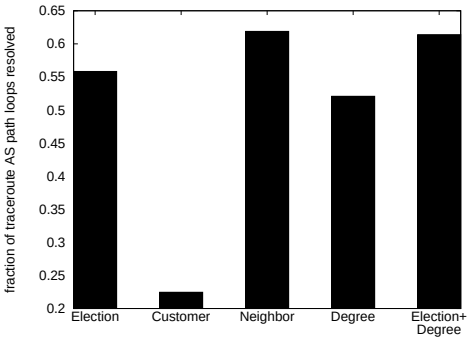

Fig. 5: Fractions of traceroute loops resolved by each heuristic.

4 Related Work

There has been signifi cant interest in studying structural properties of the Internet at the router and AS-levels for over a decade [12]. Several measurement studies have since highlighted the incompleteness of topologies inferred from publicly available routing data [9, 16, 22, 24, 32]. Much work has gone toward capturing as much of the Internet's AS-level topology as possible, most notably Zhang *et al.* [32] and He *et al.* [16]. Several large-scale active measurement projects, including Ark [17], iplane [20], and DIMES [27], use traceroutes from distributed vantage points to a large set of destinations across the IPv4 Internet. The resulting datasets have been used to reconstruct router and AS-level topologies, but merging the two views has received less attention.

A major challenge in deriving topologies from traceroute measurements is *alias resolution*, i.e., determining which interfaces belong to the same router. Tangmunarunkit *et al.* [13] proposed Mercator, a tool that attempted alias resolution by observing response packets sent from different interfaces than those probed. Spring *et al.* [29] used Ally to detect when two candidate interfaces likely shared the IP ID counter. Follow up work on alias resolution [8, 15, 28] used techniques such as IP ID counter velocities, DNS hostname conventions, and bi-directional traceroutes. Keys [19] recently documented CAIDA's attempt to expand and combine these techniques into a unified system.

There has been relatively little work on assigning routers (inferred by the previous alias resolution techniques) to the ASes that own those routers. Tangmunarunkit *et al.* [31] used a simple heuristic based on longest prefix matching to assign routers (inferred using Mercator) to ASes. Due to a lack of ground truth data, they were not able to validate their router-to-AS assignment heuristic. Tangmunarunkit *et al.* [30] was the first to study the properties of ASes in terms of the number of routers per AS. They found that ASes show high variability in the number of routers, and the number of routers per AS is highly correlated with BGP AS degree. Our work on improving AS-traceroute is inspired by the work of Mao *et al.* [23], who studied the discrepancies between traceroute-derived AS paths and BGP AS paths, and Hyun *et al.* [18], who measured the presence of third-party addresses in traceroute paths.

5 Conclusions

We have presented an approach to merge router and AS-level views of the Internet, creating a *dual graph of the Internet*. We proposed new heuristics for assigning routers from traceroute-derived graphs to ASes. We validated the success rates of our heuristics against ground truth data from a set of commercial ISPs and research networks. For multi-AS routers, the most successful heuristic was a combination of **Election** (assign the router to the AS with the largest number of interfaces) followed by **Degree** (assign the router to the AS with the smallest degree), with a success rate of 80%. For 32% of inferred single-AS routers, we either missed or mis-resolved some interface that belonged to the true owning AS, reducing our overall AS assignment accuracy to 71%. We also showed how our AS assignment techniques could be used to quantify statistical properties of ASes, as well as to improve on current state-of-the-art AS-traceroute techniques, resolving up to 62% of false loops observed in traceroute-derived AS paths.

References

1. Canet4 topology data. `http://dooka.canet4.net/`.
2. Geant topology data. `http://stats.geant2.net/lg/`.
3. I-light topology data. `http://routerproxy.grnoc.iu.edu/ilight/` .
4. Internet2 topology data. `http://vn.grnoc.iu.edu/Internet2`.
5. National lambdarail topology data. `http://routerproxy.grnoc.iu.edu/nlr2/`.
6. B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, and M. Latapy. Avoiding Traceroute Anonmalies with Paris Traceroute. In *Proc. Internet Measurement Conference (IMC)*, 2006.

7. B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: Mapped? In *Proc. Internet Measurement Conference (IMC)*, 2009.
8. A. Bender, R. Sherwood, and N. Spring. Fixing Ally's Growing Pains with Velocity Modelling. In *Proc. Internet Measurement Conference (IMC)*, 2008.
9. R. Cohen and D. Raz. The Internet Dark Matter - On the Missing Links in the AS Connectivity Map. In *Proc. IEEE Infocom*, 2006.
10. X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. Claffy, and G. Riley. AS Relationships: Inference and Validation. In *ACM SIGCOMM CCR*, 2007.
11. X. Dimitropoulos, D. Krioukov, A. Vahdat, and G. Riley. Graph annotations in Modeling Complex Network Topologies. *ACM Transactions on Modeling and Computer Simulation*, 19(4), 2009.
12. M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-law Relationships of the Internet Topology. In *Proc. ACM SIGCOMM*, 1999.
13. R. Govindan and H. Tangmunarunkit. Heuristics for Internet Map Discovery. In *Proc. IEEE INFOCOM*, 2000.
14. M. H. Gunes. APAR tool. `http://itom.utdallas.edu/data/APAR.tar.gz` (accessed 2008-07-02).
15. M. H. Gunes and K. Sarac. Analytical IP Alias Resolution. In *Proc. IEEE International Conference on Communications (ICC)*, 2006.
16. Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy. A Systematic Framework for Unearthing the Missing Links: Measurements and Impact. In *Proc. USENIX/SIGCOMM NSDI*, 2007.
17. Y. Hyun. Archipelago Infrastructure. `http://www.caida.org/projects/ark/`.
18. Y. Hyun, A. Broido, and K. Claffy. On Third-party Addresses in Traceroute Paths. In *Proc. Passive and Active Measurement Conference (PAM)*, 2003.
19. K. Keys. Internet-Scale IP Alias Resolution Techniques. *ACM SIGCOMM CCR*, 2010.
20. H. V. Madhyastha, E. Katz-Bassett, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An Information Plane for Distributed Services. In *Proc. USENIX OSDI*, 2006.
21. P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat. Orbis: Rescaling Degree Correlations to Generate Annotated Internet Topologies. In *Proc. ACM SIGCOMM*, 2007.
22. P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. Claffy, and A. Vahdat. The Internet AS-Level Topology: Three Data Sources and One Definitive Metric. *ACM SIGCOMM CCR*, 2005.
23. Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz. Towards an Accurate AS-level Traceroute Tool. In *Proc. ACM SIGCOMM*, 2003.
24. R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. In Search of the Elusive Ground Truth: The Internet's AS-level Connectivity Structure. In *Proc. ACM SIGMETRICS*, 2008.
25. RIPE NCC. Rcc12 bgp collector. http://www.ripe.net/projects/ris/rawdata.html.
26. University of Oregon RouteViews Project. `http://www.routeviews.org/`.
27. Y. Shavitt and E. Shir. DIMES: Let the Internet Measure Itself. *ACM SIGCOMM CCR*, Oct. 2005.
28. N. Spring, M. Dontcheva, M. Rodrig, and D. Wetherall. How to Resolve IP Aliases. *Technical Report UW-CSE-TR 04-05-04*, 2004.
29. N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP Topologies with Rocketfuel. In *Proc. ACM SIGCOMM*, 2002.
30. H. Tangmunarunkit, J. Doyle, R. Govindan, W. Willinger, S. Jamin, and S. Shenker. Does AS Size Determine Degree in AS Topology? *ACM SIGCOMM CCR*, 2001.
31. H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin. The Impact of Routing Policy on Internet Paths. In *Proc. IEEE INFOCOM*, 2001.
32. B. Zhang, R. Liu, D. Massey, and L. Zhang. Collecting the Internet AS-level Topology. *ACM SIGCOMM CCR*, 2005.