Towards an AS-to-Organization Map*

Xue Cai[1] John Heidemann[1] Balachander Krishnamurthy[2] Walter Willinger[2]
[1] USC/ISI, Marina del Rey, CA [2] AT&T Labs Research, Florham Park, NJ
{xuecai,johnh}@isi.edu, {bala,walter}@research.att.com

ABSTRACT

An understanding of Internet topology is central to answer various questions ranging from network resilience to peer selection or data center location. While much of prior work has examined AS-level connectivity, meaningful and relevant results from such an abstract view of Internet topology have been limited. For one, semantically, AS relationships capture business relationships and not physical connectivity. Additionally, many organizations often use multiple ASes, either to implement different routing policies, or as legacies from mergers and acquisitions. In this paper, we move beyond the traditional AS graph view of the Internet to define the problem of *AS-to-organization mapping*. We describe our initial steps at automating the capture of the rich semantics inherent in the AS-level ecosystem where routing and connectivity intersect with organizations. We discuss preliminary methods that identify multi-AS organizations from WHOIS data and illustrate the challenges posed by the quality of the available data and the complexity of real-world organizational relationships.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Network topology*; C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network management*

General Terms

Measurement

Keywords

Autonomous System (AS), organization, mapping, clustering, WHOIS, Regional Internet Registry (RIR)

1. INTRODUCTION

The Internet is of great importance to millions of people and businesses today, as a source of information, entertainment, and commerce. Thus resilience of the Internet to various threats has been the topic of a large number of research papers [1,5,12–14]. This resilience has also been extensively discussed in the popular press, especially in cases of actual outages of popular services, such as the 2008 YouTube routing problem [16, 24], service interruptions for large user groups [29], loss of connectivity due to peering disputes [25] or routing errors [22, 23], and catastrophic events [4, 9, 28].

To study Internet resilience, many analyses focus on the Internet's *AS-level topology*. Autonomous systems (ASes) appear in BGP Internet routing [21] and are defined to represent a network or group of networks that operate with a common routing policy [10]. More than 30,000 ASes in the Internet today form a logical fabric that reflects the policy and business relationships necessary to manage the flow of Internet traffic. ASes and their relationships are an attractive target for analysis because paths through the AS topology are often visible in public routing tables, implying that the resulting data can be readily mined to obtain the Internet *AS-graph* [6]. Given the apparent ease with which the Internet AS-graph can be obtained, it is not surprising that it features prominently in studies of Internet resilience since Albert et al. [1]. These studies typically treat the Internet AS topology as an abstract graph and change it, perhaps by deleting nodes or removing edges, on the assumption that these changes predict the outcome of threats to the actual Internet.

In this paper, we provide evidence that the relationship between the AS graph and the organizations that make up the Internet is much richer than previously thought, making it difficult to justify conclusions based only on modifications to the AS graph. Abstracting the Internet to a simple graph and treating it as a collection of generic nodes and links ignores much of the rich semantic content inherent in the key components that make up the AS-level Internet. This semantic content is critical to its understanding and its proper use in experiments. Another basic problem that has been gradually recognized is that obtaining a reasonably accurate AS-graph from available measurements is quite challenging. Recent work has suggested that presently available AS graphs are of questionable quality [27], limiting their

*Xue Cai and John Heidemann are partially supported by the US DHS, contract number NBCHC080035, and by the NSF, grant number CNS-0626696. The conclusions of this work are those of the authors and do not necessarily reflect the views of DHS or NSF.
This work was begun when Xue Cai was an intern at AT&T Labs-Research.

IMC'10, November 1–3, 2010, Melbourne, Australia.

use in careful studies of many Internet connectivity-related problems.

Motivated by these questions about current approaches, the first contribution of this paper (Section 2) is to move beyond the traditional AS-graph view, towards a structure we call *AS-level ecosystem*. This structure aims to capture the organizations and their ASes that make up the Internet, along with the realities of router-level connectivity in Internet Exchange Points (IXPs). Consider the frequently raised questions such as: what damage to the Internet results from the outage of a particular IXP, or legal action in a particular country, or a business dispute between organizations? Answering these requires a deeper understanding of Internet semantics than simply evaluating deletion of a node or edge in an AS graph, because such prior work omits the fact that many important organizations operate multiple ASes. To our knowledge, the only prior recognition of this fact was in differences between AS paths derived from traceroute and BGP routing [11], and PCH's manually generated AS/organization directory for network operator assistance [19].

In Section 3 we present our second contribution, a number of automated methods for extracting organization-level relationships from information in the Regional Internet Registry (RIR) WHOIS databases. This work builds on prior work [11] to provide a step towards automatic identification of the mapping of ASes to the organization to which they belong.

A final, more indirect, contribution of our work results from the validation efforts (Section 4). As part of evaluating our methods, we also characterize the quality and timeliness of the WHOIS data itself, suggesting opportunities for both operators, as contributors to and users of this data, and researchers, as users of the data, to improve the data for their mutual benefits.

2. THE INTERNET AS ECOSYSTEM

Our goal is to move towards models of the Internet AS ecosystem, so we consider ASes and organizations, how they connect, and how we can use this information to improve performance and evaluate risk.

2.1 ASes and Organizations

Central to our work are ASes and organizations. An *Autonomous System (AS)* is defined in RFC1930 [10] as "a connected group of one or more IP prefixes run by one or more network operators which has *a single and clearly defined routing policy*". While this definition is prescriptive, routing policies are not necessarily enforced or even written down. A practical, descriptive definition of an AS would be whatever is identified by an AS number in BGP routing messages.

A precise definition of *organization* is more elusive, since organizations are fundamentally socio-economic arrangements with many gradations of collaboration or independence. It is difficult to adopt a formal definition that is neither too loose, joining what should be logically separate, nor too strict, separating business units or subsidiaries that should be joined. Here we define *an organization is an entity which has control over itself and is not a subsidiary of any other organization.* (A subsidiary is majority-owned by another organization.) We adopt this definition since our emphasis is to join entities that share common, Internet-relevant business decisions.

2.2 Types of Connectivity

AS relationships can reflect a number of different contractual relationships between organizations. Common relationships are peers, customers, providers, and siblings (mutual transit providers) [8]. Single organizations can use multiple ASes to facilitate other, more complex arrangements, including creating non-uniform routing policies over one organization. Note that by their very definition, AS relationships are logical or virtual in nature and say little about the physical connectivity between the ASes in question.

This many-to-one mapping of ASes to organizations motivates an organization-level view of the AS ecosystem; another explicitly logical graph structure that is coarser-grained than the traditional AS-graph. Here nodes represent organizations and two organizations are connected if there exists an AS relationship between at least one affiliated AS in each of the two organizations. Although such an organization-level view has not attracted much prior attention in the networking literature, it is of critical importance for understanding organization-level routing.

When considering ASes by themselves, it is important to recognize that they have a rich internal structure. Depending on their size, an AS's physical infrastructure interconnects a number of different, geographically dispersed Points-of-Presence (PoPs). It is generally at these PoPs where the network connects to its customers and interconnects with other networks, either directly through Private Network Interconnects (PNIs) or via public Internet eXchange Points (IXPs). IXPs are physical infrastructure managed by third parties where members (ASes) choose to exchange traffic directly by peering with each other, rather than via upstream service providers (for a cost) [3]. It is this combination of physical, AS-level, and organization-level connectivity that makes the proposed AS ecosystem an ideal candidate for careful analyses of, for example, threats to and resilience of the Internet.

2.3 Future Applications from Understanding

An understanding of the AS ecosystem can be applied to improve performance and manage risk.

Analysis and Planning: An organization-level topology can assist in Internet analysis. For example, the accuracy of prior work in AS-peering inference [8] may be improved by considering AS/organization relationships. These peering relationships and a direct understanding of organizations can also inform what-if projections of performance and reliability and so help guide selection of additional peering agreements, or build-outs to new IXPs, or siting of content provisioning or caches.

Threat Analysis: The centrality of the Internet to everyday life raises the importance of threat assessment due to accidental or malicious damage. Goals of attackers or risks of accidents may vary from partitioning to increased latency or reduced cross-section bandwidth, or may be more subtle, such as sending unauthorized traffic (e.g., spam) or eavesdropping.

Defining threats is aided by what we believe is a *non*-threat. It is mathematically true that graphs with power-law distributions of node degree are extremely robust to random node removal, yet highly vulnerable to removal of high-degree nodes. Yet we believe application of this result to Internet AS-graphs to predict Internet stability is specious, because *there is no threat corresponding to "delete a node in*

	All	**OrgID**	**Phone**	**Email**
ARIN	21k (100%)	20K (95%)	20K (94%)	20K (93%)
RIPE	19k (100%)	11K (59%)	*unavail.*	13K (68%)
APNIC	6k (100%)	*unavail.*	4K (67%)	5K (78%)
LACNIC	1.5K (100%)	*unavail.*	*unavail.*	*unavail.*
AfriNIC	0.5K (100%)	0.4K (82%)	0.5K (96%)	*unavail.*
All	*48K (100%)*	*31K (65%)*	*25K (51%)*	*38K (79%)*

Table 1: Data availability (AS count) for three fields across the 5 RIRs.

the AS graph". Simply put, high-degree ASes cannot simply "disappear"; even in cases of major bankruptcies (such as WorldCom in 2004), the affected networks provided service throughout reorganization. Instead, we must consider models and mutations that reflect Internet semantics.

As for realizable threats to the Internet's AS ecosystem, prefix hijacking, organization disputes, and IXP outage are examples that *have actually occurred* multiple times in the past. In *route prefix hijacking*, one AS mistakenly overrides another AS's route. Prefix hijacking has been both accidental and intentional [7, 22–24]. A *de-peering* is an intentional choice by one network to refuse to route traffic from a prior peer [15, 26]. *IXP outage* refers to the fact that IXPs are physical locations and as such are vulnerable to correlated physical problems including power outages [4, 17], natural disasters [9], or human-induced problems [28]. While prefix hijacking may appear similar to "deleting an AS", a careful analysis of this as well as the other two threats shows that they are *not well described* by operations on the traditional AS graphs, but become substantially more complex problems when networks can connect in multiple locations, run their business using different ASes, or use these different ASes to interconnect at multiple IXPs.

3. DISCOVERING ORGANIZATIONS

The goals of our methodology described next is modest: we want to understand how effective simple automated clustering is at discovering an AS-to-organization mapping. Our goal here is to obtain an initial lower bound on accuracy, and to identify challenges and trade-offs for future methods. We expect that better results are possible, but will require more sophisticated information extraction and clustering methods.

3.1 Data Sources

We depend on AS registration data (i.e., WHOIS data) to discover AS-to-organization relationships. Unfortunately, this data is neither complete, nor up-to-date, nor in a common, simple format. There are five regional Internet registries [20] that provide WHOIS data, with RIPE, APNIC, LACNIC and AfriNIC relying on the Routing Policy Specification Language (RPSL) [18], and with ARIN using its own format [2]. However, given the similarity of the two formats, we can generally merge them into one.

Our work is concerned with three types of records: those corresponding to **Autonomous Systems** (ASes), **organizations** (orgs, for short), and **points-of-contact** (contact, for short). To illustrate the potential of the three types of records for performing an AS-to-organization mapping, note that ASes are identified by *ASHandle* records in ARIN and *aut-num* records in other RIRs (48K AS records, see Table 1). Some AS records are associated with an org record (via the *OrgID* or *org* fields) or administrative/technical contact information. Org records are intended to facilitate the common management of an organization's multiple records and are clearly relevant when trying to map ASes to organizations. Unfortunately, their direct use is complicated by the fact that RIR policies explicitly allow organizations to use multiple OrgID records, possibly with different OrgID's. However, org records often include administrative/technical points-of-contact information that can provide further clues about belonging to the same organization. In fact, contact records refer to individuals and can include information such as name of individual or role, a contact address, telephone numbers, and e-mail addresses. Given the role these individuals are supposed to play in managing an organization's network operations and responding to inquiries, we view telephone numbers and e-mail addresses as supporting important every-day business interactions and focus on them as providing promising clues for associating individual ASes with their parent organization.

The challenges posed by using the OrgID, telephone number, and e-mail address information buried in the AS, organization, and contact records are many. For example, only 65% of all AS records contain OrgID information, with APNIC and LACNIC providing none whatsoever; other ASes provide OrgIDs for the majority of records (59% for RIPE, 82% and 95% for AfriNIC and ARIN). However, even when coverage is good, many organizations use a number of different OrgIDs making OrgID clustering difficult. With respect to telephone numbers, they are unavailable for RIPE due to European privacy laws, and are also missing for LACNIC. Similarly, e-mail addresses are unavailable for LACNIC and AfriNIC, and partially missing for ARIN, RIPE, and APNIC (see Table 1). Clearly, the performance of any AS-to-org mapping method that relies exclusively on WHOIS data will necessarily suffer from such missing data issues and may further be impacted by the largely unknown quality of the existing WHOIS data.

3.2 Basic Clustering to Identify Organizations

Our mapping method consists of performing basic clustering on one or a number of different (canonicalized) attributes.

1. Extract (raw attributes, AS) pairs from the source data.

2. Canonicalize to (simple attribute, AS) pairs.

3. Discard *generic* attributes common to many organizations.

4. Cluster all ASes with related simple, non-generic attributes: (a) Start a new cluster c_i with some unclustered AS a_i. Then repeat until no more merges into c_i: (b) Identify c_i by the union of attributes of all its ASes, (c) Merge some other AS a_j with some matching attribute.

5. Label the cluster (Section 3.2.5).

3.2.1 *Clustering by organization IDs*

Clustering by OrgID is the simplest method and is included here as "baseline" and because it is essentially the method described in the only prior work [11] on this topic.

OrgIDs are designed to group ASes into organizations. They require no canonicalization, but we discard 8 "generic" OrgIDs belonging to 5 RIRs, IANA and 2 Network Information Centers (NIC). Given that OrgID information is not widely available and not unique (when available), we expect OrgID clustering to have poor coverage but *no* false positives.

3.2.2 *Clustering by telephone numbers*

When clustering by telephone numbers, we associate them with ASes by following three paths through WHOIS: AS-record → org-record, AS-record → contact-record, or AS-record → org-record → contact-record. We observe multiple telephone numbers with each AS (for administrative, or technical purposes), so the many-to-many telephone-to-AS relationship can cluster otherwise separate ASes.

Telephone numbers require both canonicalization (e.g., inferring country code and stripping extensions) and generic filtering. We manually build a blacklist of 11 generic telephone numbers that includes RIR, IANA, ICANN, and NIC contact numbers, as well as a few outsourcing companies. Since only 51% of ASes provide telephone numbers (see Table 1), we expect the telephone clustering method to have relatively poor performance.

3.2.3 *Clustering by e-mail domains*

To cluster based on e-mail addresses, we associate them with ASes using the same two- or three-step paths through WHOIS we relied on for telephone number-based clustering. In addition, we exploit the fact that many RIPE and APNIC records provide an additional source of administrative-pertinent e-mail addresses in the *changed* and *notify* fields. In particular, we retain the e-mail addresses in all *notify* fields and the most recent *changed* fields, providing e-mail coverage in spite of missing point-of-contact records.

Like telephone numbers, e-mail also requires careful canonicalization (e.g., discarding the user portion and keeping only the distinguishing, right-most part of the domain address) and generic filtering. Since there is no universal number of distinguishing components, we compare against a manually-built list of more than 6K suffixes with longest-suffix matching. We build a blacklist of about 50 generic e-mail domains that takes care of generic e-mail addresses in use by RIRs, NICs, outsourcing companies, and public e-mail services like Gmail and Hotmail.

While e-mail addresses capture many relationships inside organizations, their flexibility has the potential of causing many false positives. To avoid some of the more obvious misclassifications of ASes, we adjust the above basic method by implementing a simple rule designed to ensure that two clusters are not merged into one because of an e-mail address or domain used by a single individual, associated with an outsourcing effort, or affiliated with an isolated joint venture.

Given the ubiquity of e-mail in today's Internet and the critical role e-mail plays in communication between network operators within an organization or across different companies, we expect the email clustering to outperform clustering by OrgID and clustering by telephone numbers.

3.2.4 *Hybrid clustering*

While each of the above-mentioned (single-attribute) clustering methods has its strengths and weaknesses, we can combine the different types of attributes and develop two- or three-attributes clustering method to improve coverage. Our preferred method is hybrid clustering with all three attributes: OrgID, phone number, and e-mail address.

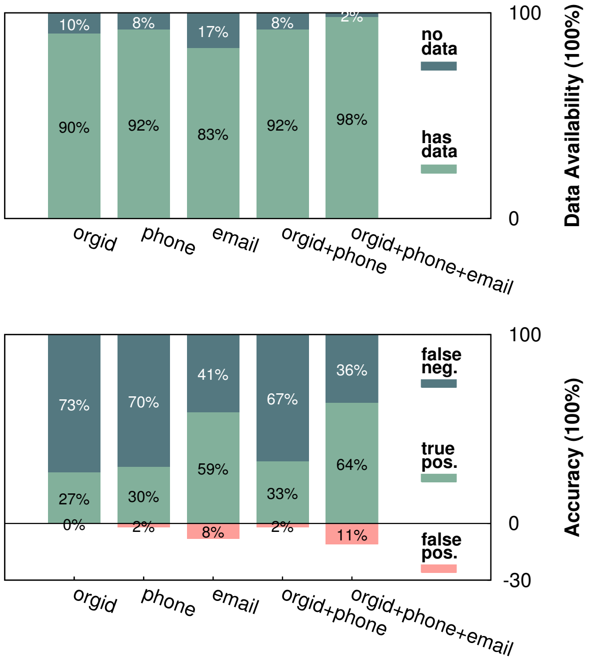

Figure 1: Validation based on real ground truth from a Tier-1 ISP, data availability (top), and biggest-cluster accuracy (bottom).

3.2.5 *Cluster Labeling and Selection*

While our basic mapping methods generate clusters, they are arbitrarily grouped, without human-friendly names. To label clusters, we extract the text names in AS and OrgID records, including *ASName*, *OrgName*, *as-name*, *descr*, and *owner* fields. Since a cluster will have many such names, often with minor variations (Org, Org Inc., Org Europe, Nippon-Org, etc.), we split these names on word boundaries to form keywords and rank keywords by their frequency in the cluster.

Finally, when comparing the accuracy of our methods against a candidate set of ASes (e.g., ground truth of all ASes belonging to an organization), we have to decide with which cluster to compare among the typically many clusters produced by a given method. For the sake of simplicity, we always compare to the *largest* cluster produced by the particular method, biasing in favor of both better coverage and larger number of false-positives (see also Section 4).

4. VALIDATION

We next evaluate accuracy and false positives and negatives of each approach. We first consider *real* ground truth from one organization's operator, then *inferred* ground truth of nine organizations as determined by manual inspection of public records.

4.1 Validation with Organization Input

We first compare our approaches against real ground truth obtained from network operators of a Tier-1 ISP. We caution that, while we have good confidence in the ground truth we were able to obtain, even operations staff suggest that this data may be incomplete.

Figure 1 shows the results of our analysis with each of our methods for the Tier-1 ISP. As a baseline, we take the ground truth of 213 ASes provided to us by the network operators. Of these 213 (100%), the top part of the figure shows that we have data of only 177 (83%) to 195 (92%) ASes for the non-hybrid methods, depending on attribute; one AS is simply missing from our bulk WHOIS data, with others omitting attributes in the data or using only generic attributes. However, by combining all three attributes together (OrgID+phone+email), the data availability rises to 98%. Thus missing or generic data leads to 8–17% of the undercount (false negatives) for the non-hybrid methods and 2%–8% for the hybrid methods.

Our automated clustering methods produce 21–50 clusters that overlap with the Tier-1 ISP, many with relevant keywords.

Selecting the largest cluster (see Section 3.2.5), we first compare the completeness (true-positive rate) of our several approaches. We see that e-mail has significantly better coverage than other approaches, finding 71% of ASes with records and 59% of ground truth, compared to 30% for telephone number and 27% for OrgID. E-mail provides better coverage because, unlike OrgID and phone numbers, e-mail addresses for different contacts can be loosely matched to cluster disparate ASes.

We next consider false positives: ASes mis-matched into the Tier-1 ISP. While our e-mail method gives better coverage, it also allows false positives. Examination of the AS record details shows 3 due to stale or incorrect data, 7 due to technical setup help, 5 due to former customer relationship, and 3 wrongly clustered because they share related attributes with 2 of these 5 ASes.

The main reason for false positives is therefore the *spectrum of relationships* between organizations—with technical support, outsourcing and joint ventures, there often is no clean organizational boundary. In 7 of 18 false positives due to e-mail clustering, we see examples of ASes acquired for customers. A secondary reason is outdated information in WHOIS, accounting for 3 of the 18 e-mail false positives.

Finally, we examine why we miss clusters (false negatives). For e-mail, 17% is due to missing or generic data, and most of these (30 of 36) are registration records using generic e-mail addresses. Clustering is *difficult when operational and customer e-mail use the same domain.* For the 52 with WHOIS data that do link to the main cluster, the dominant cause (35 of 52) is due to mergers where all WHOIS data remains in the original company's identity, with a few (5) due to independent business units not obviously linked to the parent, and some more (9) with outdated records. Three records indicate outsourced WHOIS handling, which may be related to agreements made prior to organizational changes.

Overall, we find the main clustering problem is due to registration with generic e-mail addresses and mergers that maintain all aspects of original identity, and apparently independent business units, together accounting for 33% of our missing ground truth (70 of 213) and 80% of what is missing from our cluster.

We speculate that loose matching on the names of ASes may help correct these errors. Technical support, outsourcing, both as a service offered by the Tier-1 ISP and by acquisitions cause many false positives (7 of 18) and a few false negatives (3 of 213). Stale/incorrect data and organizational churn account for a handful of the remaining (3 and 9 false positives and negatives).

Hybrid methods combine different types of attributes together to achieve a higher true-positive rate with the potential sacrifice of false-positive rate. We first start from the OrgID method with 27% coverage, then gradually add phone and email attributes. The coverage rises from 27% to 33% then to 64%, but also with a 2% and 11% increase on false-positive rate.

4.2 Validation with External Observation

We next compare our methods against manually collected estimates of clusters for nine different organizations (each with multiple ASes): four telecommunications companies, Verizon (234 ASes), Comcast (48), Time Warner (35), and China Mobile (CN Mobile) (10); four content providers, Yahoo (76), Akamai (32), Google (21), and Limelight (11); and a root-DNS provider, Internet Systems Consortium (ISC, 55). Unlike Section 4.1, the ground truth here is weaker: for each of the organizations we did our best to determine their ASes based on examination of public documents, routing data, and WHOIS data. Given the difficultly in getting strong ground truth even with operator cooperation, we know of no better way to confirm our results.

We first consider data availability (Figure 2, top). Just as the Tier-1 ISP was missing 8–17% of each kind of attribute, we see that most cases are missing up to 30% of attributes. In about half the cases (CN Mobile, Yahoo, Akamai, Limelight, ISC), many or all OrgIDs are missing; this deficit is due to poor or no OrgID coverage in APNIC/LACNIC and RIPE. We conclude that use of multiple attributes is important to provide good coverage.

Turning to coverage (Figure 2, bottom), we see that in most cases, coverage (true positives) is as good or better for these organizations compared to the Tier-1 ISP, particularly for e-mail and hybrid methods. However, for Google and Limelight, they are consistently worse, and OrgID coverage is very bad for these and Verizon, CN Mobile, and Yahoo. Coverage problems with OrgID are mainly due to attribute availability, and because some organizations such as Verizon use many (48) OrgIDs. Limelight's 11 inferred ground truth ASes break into two parts, one containing its 7 North American ASes and the other containing its 4 Asian ASes, each with separate contact information and so resulting in low coverage by all methods. Finally, Google is an unusual case because, while it has acquired other companies like YouTube, DoubleClick, GrandCentral and Picnik, to date it has done little consolidation of its RIR data.

We see a wider range of false positives for these organizations. Several show none (CN Mobile, Comcast, Google, Yahoo, Akamai) or a few, similar to the Tier-1 ISP (Verizon, TW Cable, Limelight). However, ISC shows a large number of false positives with phone or phone-related hybrid clustering. For ISC, these errors are due to outsourcing arrangements by individuals at ISC with other organizations.

Overall, we find this broader study confirms our main observations seen in the Tier-1 ISP, although they show a broader range of reactions.

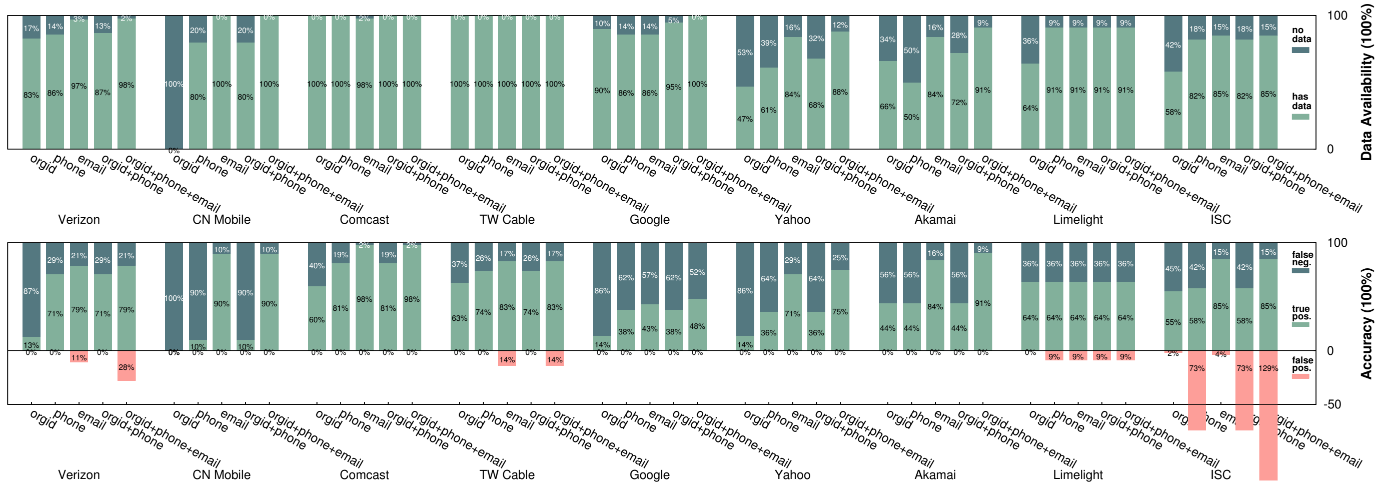

Figure 2: Validation based on inferred ground truth of 9 organizations data coverage (top) accuracy (bottom). Comparison between inferred ground truth and the biggest cluster by each method.

5. CONCLUSIONS

This paper has defined the problem and challenges of associating ASes with organizations and has described our first approach that uses WHOIS data from the RIRs. Although the WHOIS database has incorrect, missing, and outdated entries, it also has valuable information. However, relying on this data turns tasks such as clustering ASes by organization into hard problems. While the ad-hoc clustering method described in this paper shows promise, a proper AS clustering mechanism will (i) require incorporating methods for giving higher importance to attributes that are more critical to the business of running an AS, and (ii) necessitate ways to bring in additional ambient information for associating existing ASes with their parent organization. Developing a more *informed* AS clustering algorithm and running it on the nearly 30,000 ASes in the Internet is part of our future work and we hope it will lay the groundwork to better understand the characterisitcs of the Internet ecosystem.

Acknowledgments

We thank all those who contributed valuable information regarding ASes in the Tier-1 ISP. We also thank Myra Agostino for her considerable help in answering our questions.

6. REFERENCES

[1] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance in complex networks. *Nature*, 406:378–382, July 27 2000.

[2] ARIN. Introduction to ARIN's database. https://www.arin.net/knowledge/database.html, Mar. 2010.

[3] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: mapped? In *IMC '09: Proceedings of the 9th ACM Internet measurement conference*, pages 336–349, New York, NY, USA, 2009. ACM.

[4] J. H. Cowie, A. T. Ogielski, B. Premore, E. A. Smith, and T. Underwood. Impact of the 2003 Blackouts on Internet Communications. https://www.renesys.com/tech/presentations/pdf/Renesys_BlackoutReport.pdf, Nov. 2003.

[5] D. Dolev, S. Jamin, O. Mokryn, and Y. Shavitt. Internet resiliency to attacks and failures under BGP policy routing. *Computer Networks*, 50(16):3183–3196, 2006.

[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In *Proc. of ACM SIGCOMM*, pages 251–262, Cambridge, MA, Sept. 1999.

[7] A. Freedman. 7007: From the horse's mouth. NANOG mailing list, Apr. 1997.

[8] L. Gao. On inferring autonomous system relationships in the Internet. *ACM/IEEE Transactions on Networking*, 9(6):733–745, Dec. 2001.

[9] D. Greenlees and W. Arnold. Asia scrambles to restore communications after quake. International Herald Tribune, http://www.nytimes.com/2006/12/28/business/worldbusiness/28iht-connect.4042439.html?_r=1, Dec. 2006.

[10] J. Hawkinson and T. Bates. Guidelines for creation, selection, and registration of an autonomous system (AS). RFC 1930, Internet Request For Comments, Mar. 1996.

[11] Y. Hyun, A. Broido, and k. c. claffy. Traceroute and BGP AS path incongruities. Technical report, UCSD CAIDA, 2003. Published as web page http://www.caida.org/publications/papers/2003/ASP/.

[12] Z. M. J. Wu, Y. Zhang and K. Shin. Internet routing resilience to failures: Analysis and implications. In *Proc. ACM CoNEXT'07*, New York, US, 2007.

[13] A.-L. B. M. Newman and D. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[14] R. N. M. Omer and A. Mostashari. Measuring the resilience of the global Internet infrastructure system. In *Proc. of the 2009 IEEE International Systems Conference*, Vancouver, Canada, 2009.

[15] A. P. Martin Brown and E. Zmijewski. Peering Wars: Lessons Learned from the Cogent-Telia De-peering. http://www.renesys.com/tech/presentations/pdf/nanog43-peeringwars.pdf, June 2008.

[16] D. McPherson. Internet routing insecurity: Pakistan nukes YouTube? Security to the Core: the Arbor Networks Security Blog, Feb. 2008.

[17] R. Miller. Car crash triggers Amazon power outage. *DataCenter Knowledge*, May 2010. http://www.datacenterknowledge.com/archives/2010/05/13/car-crash-triggers-amazon-power-outage/.

[18] R. NCC. Ripe database query reference manual. http://www.ripe.net/db/support/query-reference-manual.pdf, Nov. 2009.

[19] Packet Clearing House. PCH INOC-DBA, Apr. 2010.

[20] Regional Internet Registry. http://www.afrinic.net/, http://www.apnic.net/, http://www.arin.net/, http://www.lacnic.net/, http://www.ripe.net/, Nov. 2009.

[21] Y. Rekhter and T. Li. A border gateway protocol 4 (BGP-4). RFC 1771, Internet Request For Comments, Mar. 1995.

[22] Renesys. Internet-Wide Catastrophe–Last Year. http://www.renesys.com/blog/2005/12/internetwide_nearcatastrophela.shtml, Dec. 2005.

[23] Renesys. Con-Ed Steals the 'Net. http://www.renesys.com/blog/2006/01/coned_steals_the_net.shtml, Jan. 2006.

[24] Renesys. Pakistan hijacks YouTube. http://www.renesys.com/blog/2008/02/pakistan_hijacks_youtube_1.shtml, Feb. 2008.

[25] Renesys Blog. Sprint and Cogent Peer, Nov. 2006.

[26] M. Ricknas. Sprint-Cogent Dispute Puts Small Rip in Fabric of Internet. http://www.pcworld.com/businesscenter/article/153123/sprintcogent_dispute_puts_small_rip_in_fabric_of_internet.html, Oct. 2008.

[27] M. Roughan, J. Tuke, and O. Maennel. Bigfoot, Sasquatch, the Yeti and other missing links: what we don't know about the AS graph. In *Proc. of 8th ACM Internet Measurement Conf.*, pages 325–330, Vouliagmeni, Greece, Oct. 2008. ACM.

[28] The National Academic Press. The Internet under crisis conditions: Learning from the impact of September 11. http://books.nap.edu/openbook.php?isbn=0309087023, 2003.

[29] WIRED. Oops! AT&T Blackhole Was 4Chan's Fault, July 2009.