Eyeball ASes: From Geography to Connectivity

Amir H. Rasti
University of Oregon
amir@cs.uoregon.edu

Nazanin Magharei
University of Oregon
nazanin@cs.uoregon.edu

Reza Rejaie
University of Oregon
reza@cs.uoregon.edu

Walter Willinger
AT&T Labs—Research
walter@research.att.com

ABSTRACT

This paper presents a new approach to determine the geographical footprint of individual Autonomous Systems that directly provide service to end-users, *i.e.*,*eyeball ASes*. The key idea is to leverage the geo-location of end-users associated with an eyeball AS to identify its geographical footprint. We leverage the kernel density estimation method to estimate the density of users across individual eyeball ASes. This method enables us to cope with the potential error associated with the location of individual end-users while controlling the level of aggregation among data points to capture a geo-footprint at the desired resolution. We use the resulting geo-footprint of individual eyeball ASes to identify their likely Point-of-Presence (PoP) locations.

To demonstrate our proposed technique, we use the inferred geo-locations of 48 million users from three popular P2P applications and assess the geo- and PoP-level footprints of 1233 eyeball ASes. The validation of the identified PoP locations by our technique against online information and prior results by a commonly-used technique based on traceroute shows a very high accuracy. Leveraging the acquired PoP locations, we examine the implications of geo-footprint of eyeball ASes on their connectivity to the rest of the Internet. In particular, we present a case study that reveals a much more complex picture of AS-level connectivity as compared to what the more traditional but geography-agnostic BGP- or traceroute-based approaches depict.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Network topology, Distributed networks*

General Terms

Measurement, Algorithms

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

Keywords

Autonomous System (AS), AS geography, Point-of-Presence (PoP), Eyeball AS

1. INTRODUCTION

As a network of networks, the Internet consists of some 30,000 inter-connected Autonomous Systems (ASes). This AS-level topology has been the focus of much research in the past decade, with studies that range from measurements and inference [15] to modeling and analysis [13] and the development of synthetic topology generators [14]. In fact, much of the research in this area has been fueled by large-scale data collection projects (*e.g.*, [10, 18, 20]) that have resulted in a high volume of readily available BGP-based or traceroute-based measurements. These datasets have been used to infer the Internet's AS-level topology as a graph where nodes are ASes and edges indicate business relationships (*e.g.*, customer-provider, peer-to-peer) between ASes.

More recently, this graph view of the AS-level Internet has been questioned. First, there has been an increasing awareness that the available BGP- or traceroute-based measurements are of limited quality to obtain an accurate and complete picture of the AS-level connectivity structure of today's Internet (*e.g.*, see [1, 19] and references therein). Second, the models that this graph view has motivated are largely descriptive in nature and essentially agnostic to the main forces responsible for shaping the structure and causing the evolution of this inherently virtual rather than physical topology of the Internet.

Partly in response to this criticism, alternative approaches to study the AS-level Internet have been advocated that align more closely with the real-world business relationships and practices encountered in the logical fabric of the Internet (see for example [4] and [6] and references therein). These recent efforts often start with the realization that ASes are not generic nodes but are entire networks that operate for a purpose and have a rich internal structure. Depending on an AS's size, its network interconnects a number of geographically dispersed points-of-presence (PoPs), where it connects to its customers or interconnects with other networks, either directly or via Internet eXchange Points (IXPs). The importance of AS geography (*i.e.*, geographic coverage or reach, number and location of PoPs, presence at IXPS) is further highlighted by the fact that the peering contracts of many ASes list explicit and geography-specific requirements for potential peering partners. For example, AS X will only peer with AS Y if Y's geographic reach is sufficiently large, or X and Y have a certain number of overlapping PoP locations, or X and Y are both present at a certain number of IXPs. Unfortunately, little is known in general about the geography of most ASes, with the possible exception of their presence at IXPs that was specifically examined [1].

In this paper, we outline a promising approach to tackle the problem of AS geography; that is, inferring an AS's geographic coverage (*geo-footprint*) and identifying its likely PoP locations. Our approach is complementary to the traditional BGP- or traceroute-based method of inferring AS-level connectivity, in both perspective and type of data used. First the traditional approach is known to perform in general increasingly worse the closer to the "edge" of the network (*i.e.*, end users) the measurements are made [5]. However, our approach starts at the "edge" (*i.e.*, "eyeballs") and experiences increasingly more difficulties as we move away from the edge towards the core of the Internet. In terms of data, instead of using BGP or traceroute data, our approach relies on the geographical location of end users or "eyeballs" (*i.e.*,, IP addresses) that are associated with an AS. In particular, the starting point of our work is a dataset consisting of the IP addresses of about 48 million users of three popular P2P applications that map to a total of 1233 "eyeball" ASes. Our main contributions are:

- We present a general methodology for determining the geo-footprint of eyeball ASes by leveraging the geo-location of their end users (Sections 2, 3).
- We use the above method to identify the likely PoP locations of an eyeball AS by associating areas with high user concentration with close-by cities in its geo-footprint (Section 4). We validate our approach using published PoP data from a number of ASes that make this information available on their websites.
- Using the PoP locations identified by our method and the AS-level connectivity resulting from a state-of-the-art inference approach, we show that the world of peering relationship at the "edge" of the network is highly diverse and complex. For example, even simple eyeball ASes tend to peer very actively at local and remote IXPs, especially in Europe, and also maintain rich upstream connectivity

The question of how to leverage the geo-properties of an eyeball AS to predict likely scenarios of how the AS connects to the rest of the Internet is left for future work. In view of our preliminary findings, a major challenge will be to explain the observed rich connectivity structure of eyeball ASes and characterize it in a quantitative manner.

2. OUR APPROACH: AN OVERVIEW

The basic idea of our approach is to use the location of end-users (*i.e.*, customers) of an AS to infer the AS's geographical reach (geo-footprint) as well as its PoP locations. To achieve this goal, our method consists of the following four steps:

- *Sampling end-users*: We collect a large number of IP addresses associated with Internet users.
- *Mapping end-users to locations*: We map individual IP addresses (or users) to their geo-location.
- *Grouping end-users by AS*: We use BGP information to group users to their corresponding ASes.
- *Estimating AS geo-footprints*: We leverage the collection of geo-locations of end-users associated with an eyeball AS to determine the geo- and PoP-level footprint of that AS.

There are three main reasons for our focus on eyeball ASes. First, the geo-features and connectivity of eyeball ASes indicate how end-users connect to the rest of the Internet. These eyeball ASes are not adequately visible to traceroute- or BGP-based approaches. Second, the accuracy of IP-geo mapping tools is significantly higher for IP addresses associated with end-users compared to infrastructure nodes [22]. Lastly, it is feasible to obtain a collection of IP addresses associated with end-users from eyeball ASes. Next, we provide further details for the first three steps of our approach. The last step, estimating geo- and PoP-level footprints, is described in Sections 3 and 4.

Sampling End-users: We crawl three large-scale P2P applications (*i.e.*, Kad, BitTorrent and Gnutella) during the months of January to June of 2009 to obtain more than 89.1 million unique IP addresses associated with end-users (peers) of these applications.

Mapping Users to Locations: To estimate the geo-location of each IP address, we examined several IP geo-location tools and databases and selected *GeoIP City from Maxmind* [16] and *IP2Location DB-15 from Hexasoft* [9] because of their reputation and coverage. Each of these databases map any IP address to a geo-location record with the following format *(city, state, country, longitude, latitude)*. The resolution of the provided coordinates is zip codes in each city, *i.e.*, all users in a given zip code are mapped to the same coordinates. We eliminated roughly 2.4M peers for which at least one of the databases did not provide city-level location. Since the two IP-geo mapping databases are from independent sources, we use the difference between their reported locations for each peer as a measure of error in IP-geo mapping [1]. We use GeoIP City as the main reference for IP-geo mapping in our analysis and use IP2Location as a second reference to estimate the error in IP-geo mapping. Using this notion of error, we remove all IP addresses whose error is larger than the diameter of typical metropolitan area, around 100km. We further elaborate on the selection of this threshold in Section 3.

Grouping Users by AS: We also group the users based on their AS affiliation using archived BGP tables from the routeviews[18] database collected during the same time period that our P2P data was gathered. To ensure a minimum density of samples in each AS, we eliminate all ASes with less than 1000 peers.

Target Dataset: Conditioning our dataset based on error in geo-location and density of sampled peers per AS significantly decreases any noise that could affect our analysis. However, it also reduces the total number of peers to 48 million and the corresponding number of eligible eyeball ASes to 1233. We call this set of ASes our *target dataset*. Given the location of all peers associated with an AS, we can broadly classify all ASes in this target dataset into city-, state-, country-, continent-level, or global ASes by identifying the smallest geographical region that contains a large

[1]While this measure of error may not be accurate, it provides a first-order approximation of geo error and could be useful to conservatively remove problematic peers with potentially large error in their geo-location.

Table 1: Profile of the target eyeball ASes.

	#Peers by source(k)			#ASes by level		
Region	Kad	Gnu	BT	City	State	Country
NA	1218	8984	1761	36	162	129
EU	18004	2519	2529	60	76	292
AS	17865	1606	1016	117	35	134

majority (>95%) of the associated peers. Table 1 summarizes the number and level of our target ASes in North America(NA), Europe (EU) and Asia (AS).

3. ESTIMATING GEO-FOOTPRINT

Given the locations of peers associated with an eyeball AS, our first goal is to infer the geographical region(s) where the AS offers service to end-users (*i.e.*, its geo-footprint) and estimate the density of users throughout the identified regions. We use a Kernel Density Estimation (KDE) [2] method with a Gaussian kernel function to estimate the probability density of customer population for an eyeball AS based on the locations of peers associated with that AS. More precisely, we place a bivariate kernel function with a predefined bandwidth at the geo-location of individual users of the AS. The aggregation of these kernel functions forms a function that estimates the overall user density over the map for each AS as shown in Figure 1. The largest contour of the aggregate density represents the geo-footprint of the AS at certain levels of resolution and may consist of one or multiple partitions. The geo-footprint of an AS clearly highlights the area within a state, country, or continent where an AS offers service, and its pronounced peaks indicate the main places with high user concentration throughout the covered region. This geo-footprint provides useful information about the services offered (*e.g.*, residential vs. retail) and connectivity provided (many vs. a few peaks) by individual eyeball ASes.

The KDE method presents a weighted average across close-by peers that serves two purposes. First, averaging smooths out the effect of error in IP-geo mapping across close-by users and provides a more reliable estimation of user density. Second, averaging offers a more aggregate (lower resolution) view (city- or state-level) of the users that is typically more useful than a detailed user-level view for assessing the geo-footprint of an AS.

3.1 Setting the Kernel Bandwidth:

The level of smoothing (*i.e.*, scope of averaging) performed by the KDE method is directly controlled by the bandwidth of the kernel function. Increasing the bandwidth leads to aggregation over a larger geographical region that has two important effects. First, it results in a coarser resolution and thus less accurate estimation of the geo-footprint for an AS. In fact, the bandwidth of the kernel function can be viewed as a tuning parameter that offers a multi-resolution view of an eyeball AS's geo-footprint. Figure 1 clearly demonstrates how increasing bandwidth can change the resolution of the geo-footprint from city- to region- and finally country-level. Second, averaging smooths out the variations in user density which makes the distinction of (smaller) peaks more difficult. It is therefore desirable to set the bandwidth so that the following two conditions are satisfied: *(i)* the resulting geo-footprint should have the desired resolution, and *(ii)* the expected geo-location error across the provided users should be filtered out. In summary, the larger value of the minimum bandwidth required by each one of these conditions determines the proper bandwidth value for the kernel function. For example, samples with a large *geographical mapping error (geo error)* cannot provide reliable city-level resolution of an AS's geo-footprint.

In our analysis, we focus on the *city-level* resolution because it provides the most useful view for detecting the main concentration point of users in order to infer likely PoP locations for each eyeball AS. To achieve this goal, the bandwidth should be larger than the average radius of a city which is around 30-35km. We set the bandwidth of the kernel function to 40km to achieve aggregation over a slightly larger region and avoid multiple peaks over a single city (*e.g.*, a separate peak for each zip code).

To determine a lower bound for the bandwidth based on geo error, we could set the bandwidth for each AS to the 90th percentile of geo error across all peers in that AS. This would result in an AS-dependent bandwidth selection. Instead, we remove all the ASes whose 90th percentile of geo error is larger than 80km. This is the main justification for removing all peers with geo error larger than 80km from our initial dataset. This strategy allows us to set the bandwidth to 40km for all ASes to obtain a city-level resolution. This choice simplifies the comparison of geo-footprints across different eyeball ASes

For a different application of the KDE method to IP geolocation, see the recent paper by Eriksson *et al.* [7] that casts the problem as a machine-learning-based classification. Their approach is relevant for fusing information from different datasets so that areas with low information content from one dataset can be compensated with information from other datasets.

4. ESTIMATING POP-LEVEL FOOTPRINT

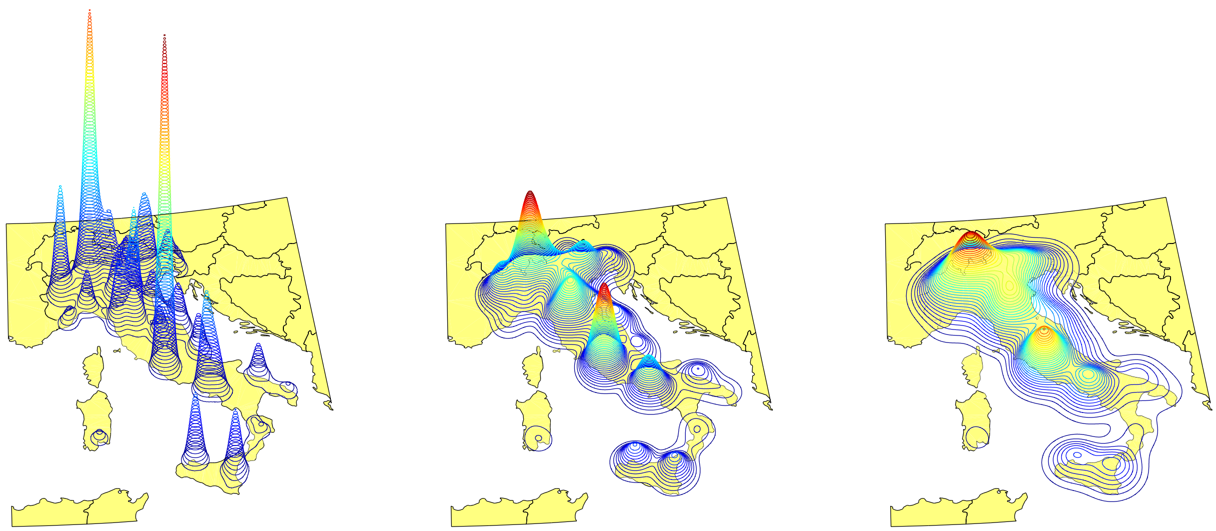
A geo-footprint of an eyeball AS can be summarized or represented by the list of major cities where significant portions of its customers are located. Intuitively, each AS must have a proportional level of presence (*e.g.*, PoP) in areas where there is a high concentration of customers. Therefore, this representation of an AS geo-footprint offers a reasonably reliable view of its PoP-level infrastructure that we call *PoP-level footprint*. Since eyeball ASes usually connect to their provider, peering and customer ASes at their PoPs, the PoP-level footprint also reveals valuable information about the location(s) of connections between related ASes.

4.1 Estimating PoP Coordinates

To extract the PoP-level footprint of an eyeball AS from its geo-footprint, we proceed as follows. First, we identify the geo-coordinates of all the local maxima $D(i)$ (*i.e.*, peaks) in the estimated density function and determine the density value of the highest peak (D_{max}). Next we select all the peaks $D(i)$ with a relatively large density compared to the highest peak, *i.e.*, ($D(i) > \alpha * D_{max}$), where α is a threshold that determines the range of density values that are considered for PoP identification. We set α to 0.01 to conservatively select peaks with a density of at least two orders of magnitude below D_{max}.

4.2 Mapping PoPs to Cities

The coordinates of identified major peaks may not directly map to a city due to the combined effects of selected



(a) Kernel bandwidth = 20km

(b) Kernel bandwidth = 40km

(c) Kernel bandwidth = 60km

Figure 1: 3D visualization of user density using KDE method for AS 3269 (in Italy) using 2.2M samples with different values for kernel bandwidth.

bandwidth, threshold for peak selection (*i.e.*, α), and the distribution of user population around each city. To address this issue, we map identified peaks to a particular city in a "loose" fashion as follows: we assume that PoPs are more likely to be located in the most populated city of a given region. For each identified peak, we examine a circular region with a radius equal to the selected kernel bandwidth around the location of the peak and map the peak to the city with the largest population in that circular region. Otherwise, we report "no city" for a peak. Using small values for α may result in an error whereby minor peaks of user density get selected at locations where a small number of users are randomly clustered due to their geo error.Using a proper α threshold, we can filter out such peaks if the selected location is not in the required vicinity of any city with sufficiently large population. The resulting PoP-level footprint obtained by this process consists of a list of cities sorted by their associated user density where PoPs of an eyeball AS are likely to be located. The user density of each PoP quantifies the level of presence of an AS in that city. For example, PoP-level footprint of AS 3269 using kernel bandwidth of 40km, as depicted in Figure 1(b), is as follows: [Milan (.130), Rome (.122), Florence (.061), Venice (.054), Naples (.051), Turin (.047), Ancona (.027), Catania (.027), Palermo (.026), Pescara (.017), Bari (.015), Catanzaro (.007), Cagliari (.005), Sassari (.001)].

4.3 Bias in Sampling Different Locations

We crawl peers in major P2P applications to sample customers of eyeball ASes from different geo-locations. Uneven penetration of P2P applications among Internet users in different ASes and locations could introduce bias to our samples. However, it is generally difficult to clearly distinguish the small market share of an AS from low penetration of a P2P application in a particular city. This potential bias can be qualitatively considered at two different levels.

1) Mild Bias: This scenario occurs when the fraction of sampled peers for an AS A in city C has a noticeable density ($D_A(C) > \alpha \cdot Max(D_A)$) but is disproportional with respect to the total number of AS customers in C. In this case, the derived PoP-level footprint of the AS includes city C as a PoP but the density value associated with C is inaccurate.
2) Significant Bias: A significant bias in collected samples could result from having a negligible (or zero) fraction of samples from a particular PoP location for a given eyeball AS. In this case, our approach does not discover that PoP location. However, for an AS with a sufficiently large number of samples, the probability of not capturing a major PoP (with a large number of customers) should be rather small. We do not examine sampling bias in this study and leave this for future work.

Another issue is whether the strategy for IP-geo mapping leverages the location of PoPs for each AS to estimate the location of end-users in that AS. In this case, our approach simply identifies the PoP locations that were used for IP-geo mapping. Our private communication with maxmind.com confirmed that the IP-geo mapping strategy relies on the information provided by users through online surveys, and information from Internet registries and ISPs. Since the actual location of PoPs for each ISP is unknown and thus not considered for determining the geo location of users in an ISP, the identified PoP locations by our approach are not affected by the mapping strategy used for each dataset.

5. EVALUATION

This section summarizes the preliminary evaluation of our proposed technique. Towards this end, we collect the reported PoP information of some eyeball ASes on the Web as the ground truth for validation. Unfortunately, collecting this information is a rather tedious task since many ISPs do not post this information online or do not use a consistent terminology or approach for listing these PoPs. For exam-

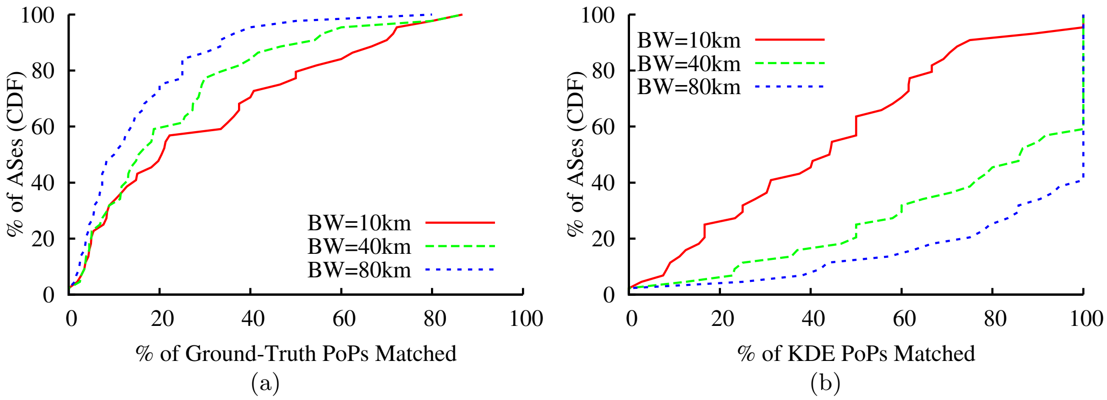


Figure 2: Validation of our technique with the reported PoP information online. (a) Percentage of ground-truth PoPs found by our method. (b) Percentage of PoPs reported by our method that match a ground-truth PoP.

ple, some ISPs may consider their access points as a PoP or list their PoPs of their peering ISPs as their own.

We focused on 672 state- or country-level ASes in our target dataset and searched the Web for their PoP information. We were able to identify PoP information for a total of 45 eyeball ASes (10 state-, 33 country-, and 2 continent-level) across North America and Europe [2]. We consider this information as ground truth and call this our reference dataset. Overall, our approach on average identified 31.9, 13.6 and 7.3 PoPs per AS with kernel bandwidth of 10km, 40km and 80km, respectively. The average number of reported PoPs per AS in our reference dataset is 43.7. We match a discovered PoP location by our technique for each AS with a reported PoP locations in the reference dataset if their relative distance is less than the radius of a city (*i.e.*, 40 km), *i.e.*, matching PoPs at the city level. Figure 2(a) depicts the distribution of the percentage of PoPs in the reference dataset that are matched with the identified PoPs by our techniques using different bandwidth values. When kernel bandwidth is 40km, for the bottom 60% of ASes, the fraction of matched PoP locations in the reference dataset is less than 20%. However, this ratio is larger than 50% for the top 10% of ASes. Furthermore, this figure suggests that using lower bandwidth generally results in mapping a larger number of PoPs in the reference dataset.

Figure 2(b) illustrates the opposite view by showing the distribution of the percentage of discovered PoP locations by our technique for each AS that match a reported PoP in the reference dataset. This figure reveals that with the bandwidth of 80km, 60% of ASes exhibit perfect match. Interestingly, decreasing the value of kernel bandwidth to 40km and 10km rapidly drops the percentage of perfect match to 41% and 5%, respectively. *Collectively, these results indicate that using larger kernel bandwidth leads to a smaller but more reliable set of PoP locations for most ASes.*

Our preliminary examination revealed that the following factors appear to cause the mismatched PoPs: First, some eyeball ASes seem to use certain PoPs in locations away from their regular customers to connect to provider (or peering) ASes. Since these ASes do not serve end-users, our approach is not able to identify them. Second, some eyeball ASes have a few PoPs within a relatively short distance. Using the KDE approach especially with moderate to large bandwidth does not distinguish these PoPs. As part of our future work, we plan to use different kernel bandwidth and determine these PoPs based on the relative distance and user density of associated peaks with different bandwidths. Third, we might have mis-interpreted a non-IP PoP as valid PoP from the obtained information online or a PoP location might be missing due to the obsolete online information. We plan to explore these issues in our future work

We have also compared our discovered PoP locations with the PoP coordinates reported in a recent traceroute-based study by the DIMES project [21]. The overlap between the two datasets consists of 226 eyeball ASes across EU and NA. While for those common eyeball ASes, our approach identified 7.14 PoPs per AS on average (with bandwidth=40km), DIMES reports only 1.54 PoPs per AS. We match a discovered PoP location by our technique for an AS with a reported PoP coordinates in the DIMES dataset within 40km distance. Our results show that for 80% of eyeball ASes our identified PoPs are a clear superset of reported PoPs by the DIMES project.

6. AS CONNECTIVITY AT THE "EDGE": A CASE STUDY

Having derived the geo-footprint and PoP locations of eyeball ASes, we next examine what this information may enable us to say about how these eyeball ASes connect to the rest of the Internet. Our comparisons are made against the current state-of-the-art "best effort" ground truth for AS-level Internet connectivity and is provided by two different datasets. For customer-provider relationships, we rely on the CAIDA AS relationships data set [3] and for peer-to-peer relationships at IXPs, we consult the dataset produced by the IXP mapping project [1].

To illustrate the challenges of making any claims about real-world AS-level connectivity at the "edge" of the network, we present a case study involving a metropolitan-area eyeball AS in Europe. Specifically, we consider AS8234 (RAI - Radiotelevisione Italiana). Based on our data, this AS has 3,000 P2P users whose geo-locations are all mapped to the

[2]This information is posted online at http://mirage.cs.uoregon.edu/AS2PoP

city of Rome. As a city-level eyeball AS, we expect it to have one or two regional or country-wide upstream providers. Examining the geo-footprints of some of our Italy-wide eyeball ASes, a natural choice of such a provider is AS1267 (Infostrada) for which we observe 1470K P2P users and obtain PoP locations across Italy, including Rome. The large number of P2P users for this ISP suggests that its major business is selling Internet connectivity to residential customers across Italy, and examining the company's website confirms this. Expecting at least one alternative connection of RAI to the rest of the Internet, plausible options include another upstream provider (possibly with more global reach than Infostrada) or peering at the Rome IXP NaMEX with a selected number of tier-2 ISPs.

However, when comparing against the best effort ground truth which we validated by performing a set of selective traceroute experiments, we encounter a substantially more complex AS-level connectivity picture for RAI. For one, this Rome-based eyeball AS has a total of five upstream providers: Infostrada (as expected) and Fastweb, two Italy-wide ISPs; Easynet and Colt, two service providers with global reach; and BT-Italia, Italy's legacy ISP. Moreover, while RAI is not present at the Rome IXP, it is a member of the Milan IXP MIX and peers there with three other ASes (*i.e.*, GARR - the Italian academic and research network, ASDASD - an Italian network provider, and ITGate - an Italian Internet service company). The two unexpected findings are the richness of upstream connectivity of this eyeball AS and its decision to peer remotely at MIX rather than locally at NaMEX.

Thus, when trying to determine the actual upstream connectivity of eyeball ASes such as RAI, one may quickly run into a bewildering web of real-world peering relationships [17, 8, 12, 11]. In some cases, a partial explanation of this richness in AS connectivity may be the separate treatment of residential and business customer traffic; *e.g.*, residential traffic is carried by one upstream provider, while commercial traffic is sent on to a different provider. In the case of RAI, having the legacy ISP BT-Italia as an additional provider may be more a historical artifact than a strategic business decision. Dual connectivity to upstream providers with global reach may again be a strategic decision based on the eyeball's business model. With respect to RAI's remote peering at MIX, it is worth pointing out that while one of its peering partners there (*i.e.*, GARR) is also present at the Rome IXP, the two other networks (*i.e.*, ASDASD and ITGate) are not members of NaMEX. This suggests that the ability for RAI to peer with the latter two networks is important enough to forgo a cheaper local solution over a more expensive remote peering arrangement.

This example of a simple eyeball AS illustrates the challenges associated with trying to leverage the geo-properties of eyeball ASes to predict and ultimately explain their connectivity to the rest of the Internet.

7. CONCLUSION

The work described in this paper has demonstrated the potential of obtaining the geographic footprints of eyeball ASes, which in turn can be used to infer infrastructure-related properties (*e.g.*, PoP locations, AS connectivity) or business-specific features (*e.g.*, serving residential vs. business customers) of these ASes. Doing so by relying solely on measurements at the "edge" of the Internet (*i.e.*, eyeball IP addresses) provides a complementary approach to the more traditional methods for studying the AS-level Internet that exploit exclusively BGP- or traceroute-based measurements. It also suggests a possible fusion of the two approaches whereby the former is augmented with tracerouting capabilities **from** the "edge" and the latter is empowered with performing targeted tracerouting **towards** the edge of the Internet (*i.e.*, eyeballs). Such a combined approach holds the promise to unearth much of what has remained invisible in the AS-level Internet and reveal a maze of real-world peering relationships whose solution will require substantial future research efforts.

Acknowledgments

The authors are grateful to Matthew Roughan for his invaluable contributions to this project and to the anonymous reviewers for their helpful feedback. Kad and BitTorrent datasets are kindly provided by Ghulam Memon and Ruben Cuevas. We acknowledge Kaveh Kazemi for extracting PoP information for a large number of ASes from the web. We also thank Maxmind and IP2Location for providing their geo-IP mapping databases. This material is based upon work supported in part by the NSF Award Nets-NBD-0627202.

8. REFERENCES

[1] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: Mapped? In *ACM SIGCOMM conference on Internet measurement conference*, Nov. 2009.

[2] Z. Botev, J. Grotowski, and D. Kroese. Kernel Density Estimation via Diffusion. *Annals of Statistics*, 2010.

[3] The Cooperative Association for Internet Data Analysis (CAIDA). AS relationships. *http://www.caida.org/data/active/as-relationships*.

[4] H. Chang, S. Jamin, and W. Willinger. Internet connectivity at the as-level: An optimization-driven modeling approach. In *ACM SIGCOMM MoMeTools Workshop*, Aug. 2003.

[5] K. Chen, D. Choffnes, R. Potharaju, Y. Chen, F. Bustamante, D. Pei, and Y. Zhao. Where the Sidewalk Ends: Extending the Internet AS Graph Using Traceroutes From P2P Users. In *ACM CoNEXT*, Dec. 2009.

[6] A. Dhamdhere and C. Dovrolis. An agent-based model for the evolution of the internet ecosystem. In *COMmunication Systems And NETworks*, July 2009.

[7] B. Eriksson, P. Barford, J. Sommers and R. Nowak. A Learning-based Approach to IP Geolocation. In *Passive and Active Measurement conference*, Apr. 2010.

[8] P. Faratin, D. Clark, P. Gilmore, and A. Berger. Complexity of internet interconnections: Technology, incentives and implications for policy. In *Telecommunications Policy Research Conference*, Sept. 2007.

[9] Hexasoft. IP2Location. *http://www.ip2location.com/aboutus.aspx*.

[10] Y. Hyun. Archipelago Infrastructure. *http://www.caida.org/projects/ark*.

[11] R. Ma, D. Chiu, J. Lui, V. Misra, and D. Rubenstein. Interconnecting eyeballs to content: a shapley value perspective on isp peering and settlement. In *ACM*