Using PeeringDB to Understand the Peering Ecosystem[*]

Aemen Lodhi
School of Computer Science
Georgia Institute of
Technology
aemen.lodhi@gatech.edu

Natalie Larson
CAIDA
University of California
San Diego
nllarson@caida.org

Amogh Dhamdhere
CAIDA
University of California
San Diego
amogh@caida.org

Constantine Dovrolis
School of Computer Science
Georgia Institute of
Technology
constantine@gatech.edu

kc claffy
CAIDA
University of California
San Diego
kc@caida.org

ABSTRACT

In this study we mine one of the few sources of public data available about the interdomain peering ecosytem: PeeringDB [1], an online database where participating networks contribute information about their peering policies, traffic volumes and presence at various geographic locations. Although established to support the practical needs of operators, this data also provides a valuable source of information to researchers. Using BGP data to cross-validate three years of PeeringDB snapshots, we find that PeeringDB membership is reasonably representative of the Internet's transit, content, and access providers in terms of business types and geography of participants, and PeeringDB data is generally up-to-date. We find strong correlations among different measures of network size – BGP-advertised address space, PeeringDB-reported traffic volume and presence at peering facilities, and between these size measures and advertised peering policies.

1. INTRODUCTION

Although most of society considers the Internet as a critical infrastructure by now, we still know surprisingly little about its dynamics and structure. Its opaqueness is due to both the complexity of network interactions and the proprietary treatment of many aspects of these interactions by commercial providers. One source of public data on some of these network interactions that has yet to be systematically mined by researchers is PeeringDB [1]. PeeringDB [1] is an online open database where the operators of Autonomous Systems (ASes) provide information about the networks, such as peering policies, traffic volumes and presence at various geographic locations. PeeringDB was established in 2004 to assist peering coordinators identifying potential peers and peering locations. Over the last 3 years it has grown by 74% from 1950 participants in August 2010 to 3392 in August 2013.

Networks registered in PeeringDB self-report their business type, yielding a data set that can be used directly or to validate other AS business-type inference algorithms [2, 3, 4]. Second, networks report the set of IXPs and private peering facilites at which they are present. Third, networks self-report their general peering policy (either "Open", "Selective", or "Restrictive") and approximate traffic levels. Data on these four AS properties (business type, peering presence, traffic volume, peering policy) can help parameterize models of interdomain interconnection economics and traffic flow ([5, 6, 7, 8], among others). To the best of our knowledge, PeeringDB is the only centralized resource available to the research community that publishes such data.

We undertake a study of PeeringDB data to investigate three questions. First, since PeeringDB participation is voluntary, with no mechanism to verify the accuracy of reported information, we investigate whether the PeeringDB dataset is representative, correct, and current. We then explore PeeringDB data from the *network perspective*, focusing on the geographic expanse, traffic volume, address space and peering policies that networks advertise. Our goal is to discover correlations between these properties; the presence of strong correlations would allow us to estimate properties of networks that are otherwise difficult to obtain (e.g., approximate traffic levels) using a property we can estimate from publicly available data (e.g, size of advertised address space). Finally, we explore what historical snapshots of the PeeringDB database can tell us about the evolution of the Internet peering ecosytem.

We find that PeeringDB membership is representative of transit, content, and access provider populations, and that most networks keep their records current. The data less accurately reflects *IXP properties* such as member counts and their evolution over time, because many networks in developing regions do not participate in PeeringDB. We find strong correlations among different measures of network size – advertised address space (from BGP), traffic volume and geographic expanse (reported on PeeringdB), and between these size measures and the peering strategies that those networks use. The presence of such correlations allows us to estimate difficult-to-obtain network properties, such as traffic volume and peering policy, using parameters such as the BGP-advertised address space or geographic expanse that are easier to obtain. Using three years of historical PeeringDB snapshots, we observe the evolution of the peering ecosystem – geographic expansion by content, access, and transit networks that agrees with their published peering behavior, changes in traffic volume, and a shift towards more restrictive peering. Furthermore, we find widespread adoption of Open peering among transit providers, which is counterintuitive given that transit providers prefer other ASes as their customers instead of peers.

2. DATASETS

In this paper we analyze the latest (August 2013) snapshot from PeeringDB, which we refer to as the **Aug13-PDB** dataset. Participating networks can report a wide range of attributes that are stored as fields in its database record [1]; we use the following fields:

Business type of the network, which is one of *Network Service Provider (NSP)* , *Cable/DSL/Access Provider*, *Content Provider*, Enterprise, *EducationResearch*, or *Non-Profit*.

Approximate traffic volume that the network handles, which ranges from *0-20Mbps* to *1+Tbps*, in 14 distinct bins.

Peering strategy that the network uses: *Open*, *Selective*, or *Restrictive*. ASes advertising a *Restrictive* peering policy are generally not inclined towards peering. ASes advertising a *Selective* policy prescribe a set of criteria (overall traffic volume, traffic ratios, minimum number of geographic locations of overlap, etc.) that potential peers must meet. ASes advertising an *Open* policy are generally willing to peer with any co-located network.

The IXPs and private peering facilities where a network is present.

To examine whether PeeringDB participants are representative of the AS population, we construct an AS topology using BGP routing table dumps from Routeviews [9] and RIPE RIS [10] in the first week of August 2013. We use CAIDA's AS-relationship algorithm [11] to infer the number of customers of each AS. We use this BGP data to determine the size of the address space that each AS originates (removing double-counting due to ASes advertising overlapping prefixes). We classify ASes according to broad geographic regions using the RIR database (WHOIS) where the AS is registered: ARIN (North America), RIPE (Europe, Middle East, and the former USSR), APNIC (Asia/Pacific), AfriNIC (Africa), and LACNIC (Latin America). We refer to the dataset obtained from BGP and WHOIS information as the **Aug13-BGP** dataset.

Through private communication with the PeeringDB operators we found that they do not maintain historical snapshots of the PeeringDB data. However, they publish a nightly mysql dump of the entire database, which we have been archiving daily since July 2010. To the best of our knowledge, this is the only resource of historical peering data available to the research community. We will make this data available publicly via CAIDA's data sharing portal [12].

3. REPRESENTATIVENESS AND USABILITY OF PEERINGDB DATA

Given that PeeringDB runs on a volunteer basis, a key question is whether PeeringDB participants are representative of the general AS population, and whether the data is up-to-date and correct.

3.1 Business type representation of PeeringDB

We first study whether the business type of PeeringDB participants is representative of the entire AS population. The **Aug13-PDB** dataset contains 3392 ASes (7.5% of the number in **Aug13-BGP**), of which 31% are *Network Service Providers* (NSP) (Transit Providers), 25% *Content Providers*, 33% *Access Providers*, 4% *Enterprise Networks*, 4% *Educational/Research*, and 3% are *Non-profit organizations*. The **Aug13-BGP** dataset contained 45074 ASes, of which 4.5% were *Transit providers*, 4.5% *Content/Access/Hosting providers*, and 91% were *Enterprise Customers* according to our scheme for classifying ASes into business types [2]. Based on this public BGP data, enterprise customers are under-represented in PeeringDB as compared to transit, content and access networks.

To determine if the largest transit networks are present in PeeringDB, we use CAIDA's AS-rank, which ranks transit providers according to the number of ASes present in the provider's customer

Registry	Aug13-PDB All (%)	BGP All (%)	Aug13-PDB Non-stubs (%)	BGP Non-stubs (%)
ARIN	25.1	34.9	24.0	26.8
RIPE	53.2	44.2	53.6	49.8
APNIC	13.4	12.3	15.3	13.7
LACNIC	4.9	6.2	5.1	7.4
AFRINIC	1.7	1.5	1.7	2.1

Table 1: Geographical distribution of ASes in the **Aug13-PDB** and **Aug13-BGP** datasets. While the overall PeeringDB population is biased towards RIPE, the PeeringDB non-stub population is geographicly representative of the entire Internet.

cone [11]. We find that 93% of the top-100, 80% of the top-200 and 74% of the top-300 ASes from AS-rank were present in the **Aug13-PDB** dataset, including all known Tier-1 [13] and major Tier-2 ASes [14]. To determine whether popular content providers are present in PeeringDB, we used Alexa's ranking of major content sites in August 2013 [15] to find the ASes that host the most popular websites. 59% of ASes hosting the top-100, 39% of ASes hosting the top-500 and 38% of ASes hosting the top-1000 websites were present in **Aug13-PDB**. To determine whether popular access providers are present in PeeringDB, we crawled the tracker for the popular torrent site The Pirate Bay [16] over two weeks in July 2013, and obtained a list of IP addresses that connected to the tracker. We then mapped those IP addresses to ASes, and ranked ASes by the number of BitTorrent clients. We find that 54% of the top-100 ASes in terms of host count, 52% of the top-200, and 47% of the top-300 ASes were present in *Aug13-PDB*.

Limitations of AS representation in PeeringDB: Given that the objective of PeeringDB is to assist peering coordinators, it is likely to draw the attention of only that section of the network operator community that is interested in peering. Hence, we can expect organizations whose primary business is not Internet connectivity, e.g., education/research, retail enterprises, etc., networks with small traffic volumes, limited resources, small geographic footprint, to not appear in PeeringDB. Finally, some networks may not be willing to share information about themselves due to competitive reasons. This is evident as there were 8724 registered users in the **Aug13-PDB** dataset but only 3392 ASes that volunteered any information about themselves.

3.2 Geographical representation of PeeringDB

Our next question is whether the geographic distribution of PeeringDB participants matches that of all ASes seen in BGP. To answer this question we used WHOIS information to determine the RIR that each AS in peeringDB is registered in, and we compared that with the distribution for all ASes. The first two columns of Table 1 show the fraction of ASes in the **Aug13-PDB** and **Aug13-BGP** datasets associated with each registry. The APNIC, LACNIC, and AFRINIC registries have almost the same representation in the **Aug13-PDB** and **Aug13-BGP** datasets. RIPE, however, is over-represented and ARIN is under-represented in **Aug13-PDB**. Since PeeringDB membership is not representative of the entire AS population (most of which are stub networks [2]), we isolate the geographic distribution of *non-stub* networks in the two rightmost columns of Table 1. For this non-stub population, the representation bias towards RIPE (over ARIN) is much lower; the geographic characteristics of non-stub PeeringDB participants are thus similar to those of the larger Internet.

3.3 Freshness of PeeringDB records

Using the *last updated* field in PeeringDB records in the Aug 1,

2013 snapshot, we find that the median time since the last update was between 10-14 months for NSPs, Cable/DSL/Access providers and Content providers, and 17 months for Enterprise networks. When we considered the top-20 NSPs, top-20 Content and top-20 Access providers (ranked according to their advertised traffic volume), 70% of this set had updated their peeringDB records in the month preceding August 1, 2013. PeeringDB records thus appear to be reasonably current. PeeringDB does not incorporate topology data, which is more susceptible to frequent variation. We do not expect peering policies, geographic co-location and traffic profiles to change frequently.

3.4 Correctness of data reported in PeeringDB

Snijders [17] recently found that PeeringDB data was 99% accurate with respect to network presence at IXPs, i.e., 99% of the instances where a network reported presence at an IXP were true. To check the consistency of peering policies that networks report on PeeringDB and on their webpages, we obtained the peering policy URLs of 50 networks in PeeringDB, and compared the policy seen on their URL with the policy mentioned in the PeeringDB record. In each case, the peering policy listed on PeeringDB (Open, Selective or Restrictive) matched the peering policy at that network's policy URL. Verifying other self-reported network properties such as traffic volume is difficult; however, we are currently developing a method to compare a network's advertised peering policy with its peering behavior at various IXP route servers.

We investigated whether we could use PeeringDB to infer a specific *IXP property* – the number of members present at that IXP. For each of the top-20 IXPs for which we could find member lists online, we calculated a ratio of the number of members of the IXP inferred from PeeringDB to the number of members obtained from the IXP's webpage. If a network does not participate in PeeringDB but is present at an IXP, then that network does not appear in the member list created from PeeringDB. Consequently, for each of the top-20 IXPs, this ratio is less than 1; the median is 0.8. For some IXPs this ratio is close to 1, e.g, LINX Extreme LAN (0.99), LINX Juniper LAN (0.98), Seattle Internet Exchange (0.98); these IXPs encourage their members to join PeeringDB. For many IXPs the ratio is small, especially in developing regions, e.g., Moscow IX (0.25), PTT Sao Paolo (0.32) and Hong Kong IX (0.62). We conclude that IXP member counts from PeeringDB are a lower bound on the number of networks present at the IXP; they are not complete membership lists. Consequently, PeeringDB should not be used to estimate the size (in terms of member count), or the diversity of the participants at an IXP, unless we first verify that the member list generated from PeeringDB is close to that obtained from the IXP itself.

4. PROPERTIES OF PARTICIPANTS

We explore the use of PeeringDB to infer properties of networks that are difficult to obtain from other sources. We focus on three measures of a *network's size* – geographic expanse (the number of IXPs and private peering facilities), advertised traffic volume, and BGP-advertised IPv4 address space. Networks self-report the first two properties in PeeringDB; we obtain the size of the IPv4 address space from publicly available BGP data. Metrics of the size and geographic expanse of networks are important for developing, parametering, and evaluating models of interdomain economics and interconnection [5, 6, 7, 8]. Moreover, the presence of strong correlations between these properties would enable us to estimate a network property that is difficult to measure (e.g., traffic volume) using a property that is more readily available (advertised address space, or number of peering locations).
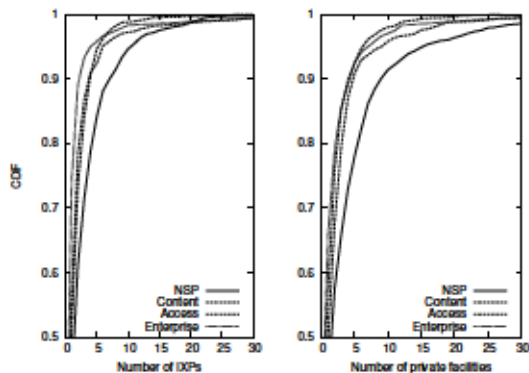

Figure 1: Distribution of the number of IXPs and private peering facilities at whcih PeeringDB participants are present. NSPs are generally present at the largest number of IXPs and private peering facilities. Enterprise networks are similar to content and access in geographic expanse.

Geographical expanse: Figure 1 shows the distribution of the number of IXPs and private peering facilities where participating networks (classified according to their self-reported business type) are present. Unsurprisingly, the self-reported data indicates that NSPs tend to have presence at more IXPs (median=2 and 90th percentile=8) and private peering facilities (median=2 and 90th percentile=9) than other network types. The median number of IXPs for Enterprise, Content and Access networks is a single IXP, while the 90th percentile is 2 for Enterprise networks and 4 for Content and Access networks. More surprising is that the presence of Enterprise networks at private peering facilities is comparable to that of content and access providers; in each category, the median is a single facility and 90th percentile is 5 facilities. Conventional wisdom suggests that enterprise networks are usually stubs at the edge of the network that do not engage in widespread peering. While the sample of Enterprise networks in PeeringDB is small (only 120 networks), and contains networks such as Amazon and Websense Hosted Security that peer at many locations, it suggests a trend toward richer peering at the periphery of the Internet.

Relation between geographic expanse and traffic volume: We examine the correlation between the geographic expanse of a network (the number of IXPs and private peering facilities) and the advertised traffic volume of that network. Figure 2 bins the total number of locations where a network is present, and shows the distribution of the traffic volume of networks in each bin. In general, the number of locations where a network is present at positively correlates with its advertised traffic volume. The fraction of networks advertising large traffic volumes (100-1000Gbps and 1Tbps+) increases with the total number of locations. The number of peering locations of a network is usually easier to discover than its traffic volume, and the correlation between these factors suggests that we may be able to roughly estimate the latter based on the former.

Relation between traffic volume and advertised address space: Figure 3 shows the median, 10th and 90th percentiles of the advertised address space size for each advertised traffic volume for different classes of networks. Access and enterprise networks have the largest median advertised address space for each traffic volume, with strong correlation (correlation coefficient 0.91), while content providers have the smallest median advertised address space for each traffic volume with weaker correlation (correlation coefficient 0.56). NSPs show the strongest correlation coefficient (0.95) be-
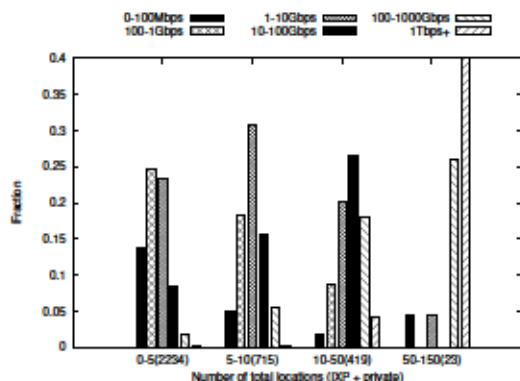
Figure 2: Number of locations at which a network is present vs. self-reported traffic volume. Networks present at more locations are more likely to advertise larger traffic volumes. Total number of networks in each category given in paranthesis.
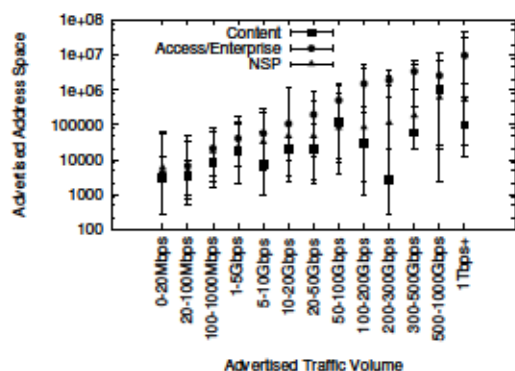

Figure 3: Reported traffic volume of a network and its advertised address space size. Networks advertising more address space also report higher traffic volumes. Access/Enterprise networks advertise more address space per unit of traffic than transit and content networks.

tween advertised traffic volume and advertised address space, but lower median values than access and enterprise networks. These trends match what we expect from these business types. Access and enterprise networks serve end-users, and their traffic volume typically increases with the size of their advertised address space; in contrast, content providers do not require much address space to serve content. Since a network's BGP-advertised address space is computable from public data, its strong correlation (for access, enterprise and transit networks) with traffic volume suggests that we may be able to use BGP data to estimate the approximate traffic volume of other networks on the Internet.

Changes in the traffic volume of participating networks: Given that reported global traffic levels in the Internet continue to increase rapidly [18], we expect that most networks should advertise larger traffic volumes over time. For PeeringDB snapshots between August 2010 and 2013, 35% of access providers, 42% of content providers, and 29% of transit networks reported a *decrease* in their traffic volume. We do not know whether this decrease is due to a loss of customers, or consolidation in the content delivery and access markets [19]. Another plausible hypothesis is that networks initially advertise inflated traffic volumes to peer with large networks. Over time, however, networks are able to determine the

actual traffic volumes being exchanged with their peers. Hence, networks advertising inflated traffic volumes may realize that doing so only leads to unstable peering relationships and drives away peers with whom they could have formed stable links. Therefore, they report figures closer to reality. Correlating observed changes in traffic volume with publicly available financial information about revenues and incomes could help identify cases where traffic volume changes are due to factors such as loss of market share (as opposed to changes due to more truthful reporting). Such *actual changes* of traffic volume may help researchers validate models that relate traffic flow to economics and strategic decisions of network.

Geographical expansion by networks: Researchers have studied the geographic expansion of networks, and the resulting *flattening* of the Internet topology [20, 6]. Historical peeringDB snapshots allow us to estimate the geographic expansion by participating networks. Of 2,525 networks present in both Aug 2010 and Aug 2013, 25% increased their presence at IXPs, and 25% increased their presence at private peering facilities. When classified by business type, 33% of NSPs present in both snapshots increased their presence at IXPs and 37% did so at private peering facilities. The increase at peering locations was 24% and 31% for Content providers, and 28% and 31% for Cable/Access/DSL providers. The following case studies from each business type illuminate the changing structure of the ecosystem.

Content providers: From 2010 to 2013, Google increased its peering presence from 57 to 72 IXPs and from 58 to 77 private facilities. Akamai's presence at private peering facilities is almost constant (35 in 2010 to 36 in 2013), while its presence at IXPs increased from 47 to 74. Limelight Networks' presence at IXPs remained constant at 42, while it expanded its presence at private facilities from 55 to 65. These observations are consistent with well-documented peering policies of these networks [21, 22], i.e., engage in Open peering at IXPs for low-traffic peers and private peering for high-traffic peers. In contrast, Limelight Networks advertises a Selective peering policy requiring a minimum of 1Gbps of traffic [23], implying that it prefers private peering with qualifying networks. The geographic expansion of Netflix follows its growth as a major source of Internet traffic. In 2010, Netflix was present at one IXP and one private peering facility; in 2013, it is present at 21 IXPs and 27 private peering facilities.

Access Providers: Major access providers, e.g., Comcast, Time Warner Cable, Vodafone and ClaraNet announce Selective or Restrictive peering policies. Vodafone and Claranet *decreased* their IXP presence between 2010 and 2013 (from 7 to 5 and 15 to 11, respectively), while Comcast did not report presence at any IXPs since 2010 (and was present at 17 private facilities in 2013). Time Warner Cable reduced its private peering locations from 12 to 10 and added a single IXP between 2010 and 2013.

Transit Providers: Large transit providers, e.g., AT&T, Level3, Global Crossing (Now Level3), TiNet, TeliaSonera, Deutsche Telekom and TATA announce Restrictive or Selective policies. AT&T and Level3 are not present at any IXP or private peering facility; presumably they prefer to peer at their own facilities. Tinet, Deutsche Telekom, TATA, and TeliaSonera have all decreased their presence at IXPs and increased their private peering count from 2010 to 2013. Hurricane Electric is an interesting exception; it advertises an Open peering policy, and has increased its IXP and private peering count (from 43 to 68 and 27 to 58, respectively).

Network presence at multi-IXP cities: PeeringDB lists 59 cities with more than one IXP. For networks in multi-IXP cities, peering at multiple IXPs could increase the diversity of peering partners and resiliency of interconnection. For each multi-IXP city with 4 or more IXPs, Figure 4 shows the fraction of networks present in
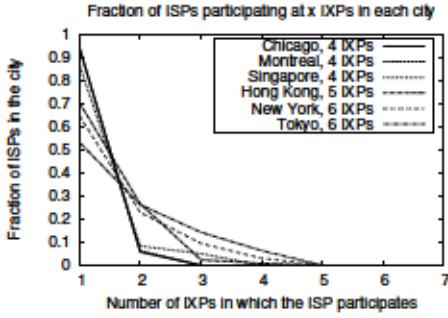
Figure 4: For each multi-IXP city with 4 or more IXPs, the fraction of networks that are present at different numbers of IXPs.

Type	Total	Open (%)	Selective (%)	Restrictive (%)
NSP	1064	66.7	28.7	4.5
Content	843	83.9	14.4	1.5
Access	1122	79.1	18.5	2.3
Enterprise	120	65.0	27.5	7.5
Edu/Research	133	69.1	28.5	2.2
Non-profit	108	81.4	14.8	3.7

Table 2: Peering strategy distribution by network type. Open peering is the dominant peering strategy irrespective of business type.
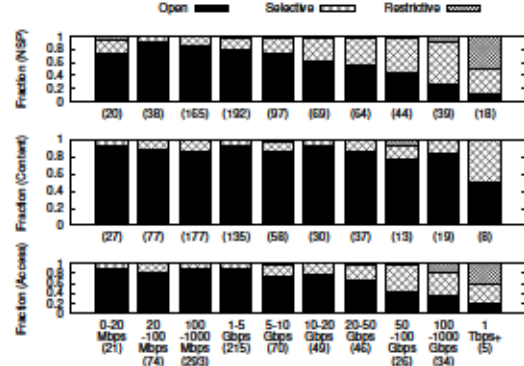

Figure 5: Peering strategy distribution of transit, content, and access providers – classification by traffic volume. Peering openness of transit and access networks is negatively correlated with geographic expanse. There is no such correlation for content providers.

that city that connect to a given number of IXPs. Perhaps unsurprisingly, there are significant differences between cities; in Chicago, Montreal, and Singapore, close to 90% of the networks are present at a single IXP in that city. London and Paris, on the other hand, present the opposite case, where 40% and 35% of participants respectively are present at 2 IXPs. Delving into why networks in different multi-IXP cities peer differently involves looking into the size, business model, and diversity of the participant mix at these different IXPs, which we plan to do in future work.

5. ADVERTISED PEERING POLICIES

Network presence at IXPs is a measure of the *ability* of networks to peer with other co-located networks, but says nothing about their *willingness* to do so. Peering policies advertised in PeeringDB can serve as a coarse measure of peering openness. We emphasize that a network is under no obligation to follow its announced peering policy. At the same time it is unlikely that networks can derive any advantage by advertising a completely different peering policy than what they follow in practice. For example, a network implementing Open policy would only drive potential peers away by advertising a Selective policy. Similarly, a network implementing Selective policy while advertising Open will form many unstable peering links as most links will fail to qualify its satisfy its peering constraints. Nevertheless, the peering policies in PeeringDB should be viewed as a coarse measure of peering openness, as many networks also require that their peers follow additional constraints, e.g., co-location at more than one IXP, traffic volume exchanged between peers, 24/7 operator support, etc. Another reason for deviation from advertised peering policies is due to the complexities of implementing import/export filters to enforce these policies when a network connects to an IXP's route server. Giotsas et al. [24] found that some networks advertising a Selective peering policy in PeeringDB were actually engaging in open peering at some IXPs, due to the complexity of setting fine-grained import/export policies at the corresponding route server.

Of the 3392 ASes in the **Aug13-PDB** dataset, 76% use *Open* peering, 21% use *Selective*, and 3% use *Restrictive*. We examine whether this preference for *Open* peering depends on other properties of these networks such as their business type, or the measures of network size (geographic expanse, approximate traffic volume).

Peering strategy distribution by business type: Table 2 shows the fraction of networks in each business type that advertise *Open*, *Selective* and *Restrictive* peering. Interestingly, the peering strategy distribution does not depend significantly on the AS business type. Between 65% to 84% of ASes from each business type advertise an *Open* peering strategy. The popularity of Open peering is coun-

terintuitive, especially for transit providers, who could mostly use *Selective* or *Restrictive* peering to increase their customer base and transit revenues. The trend towards *Open* peering is not limited to small transit providers; 32% of NSPs with traffic volume greater than 100 Gbps, 43% of providers with traffic volume between 50 and 100 Gbps, and 56% of providers that advertise a global scope use *Open* peering.

Peering strategy distribution by traffic volume: Figure 5 shows the peering strategy distribution for NSPs, content providers, and access providers that advertise a given traffic volume. For NSPs and access providers the preference for *Open* peering gradually decreases as the AS's traffic volume increases. Low volume NSPs and access providers show a strong preference for *Open* peering; 80% of 415 NSPs and 87% of 603 access providers advertising traffic volume less than 5Gbps announce *Open* peering policy. On the other hand, only 1 out of 5 NSPs and 2 out of 18 access providers advertising more than 1 Tbps of traffic declare an *Open* peering policy. Content providers have a weaker relation between traffic volume and peering policy; 88% of 573 content providers with less than 1Tbps of traffic announce *Open* peering. Of 8 content providers with more than 1Tbps of traffic, 4 announce *Open* peering, and none announce *Restrictive* peering.

Peering strategy distribution – joint classification by traffic volume and number of customers: Transit providers (and also to some extent access providers, e.g. Comcast), rely on transit customers as a source of revenue. For these providers, peering openly could mean losing revenue-generating customers. We use traffic volume and size of customer base to consider four classes of ASes:
Class-1: Small traffic volume, small number of customers (ST+SC)
Class-2: Small traffic volume, large number of customers (ST+LC)
Class-3: Large traffic volume, small number of customers (LT+SC)

Class	Total	Open (%)	Selective (%)	Restrictive(%)
Class-1 (ST+SC)	891	84.8	14.0	1.1
Class-2 (ST+LC)	19	57.9	36.8	5.3
Class-3 (LT+SC)	310	74.5	21.0	4.5
Class-4 (LT+LC)	344	46.8	45.3	7.6

Table 3: Peering strategy distribution of transit and access providers – joint classification by number of transit customers and traffic volume. Transit providers with low traffic and few customers are most likely to adopt Open peering; providers with large traffic and many customers are least likely.
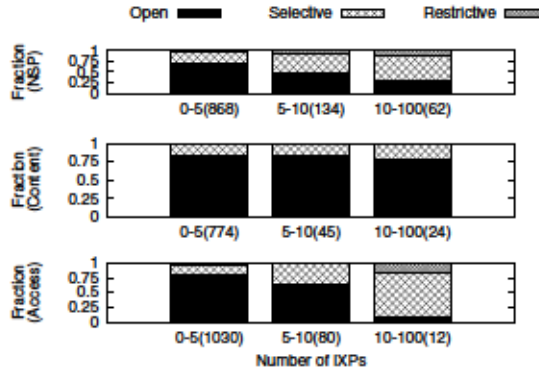

Figure 6: Peering strategy distribution of NSPs, Content, and Access providers by geographic expanse. Peering openness of NSPs and access networks is negatively correlated with traffic volume. We find no such correlation for content providers.

Class-4: Large traffic volume, large number of customers (LT+LC) Class-1 contains networks in the bottom 30% by traffic volume and bottom 30% by number of customers; Class-2 contains networks in the bottom 30% by traffic volume and top 70% by number of customers; other classes are defined similarly. Table 3 shows the peering strategy distribution for these four classes of ASes. Open peering is most common in Class-1 networks (85% of such ASes use Open peering) and is least popular for Class-4 ASes. However, 47% of even Class-4 ASes advertise *Open* peering. Using game-theoretic analysis and agent-based simulations, we have shown [25, 26] that transit providers can gravitate towards Open peering due to myopic decision making without coordination. Our results also showed that in a world with widespread Open peering, networks in Class-1 stand to gain, while networks in Class-4 stand to lose. The distribution of peering policies seen in the real world is consistent with our previous results [25, 26].

Peering strategy distribution by geographic expanse: Figure 6 shows the peering strategy distribution for NSPs, content and access providers as a function of their geographic expanse (the number of IXPs at which they are present). Similar to the earlier classification by traffic volume, the peering strategy of content providers is largely independent of geographic expanse; these ASes mostly prefer Open peering independent of their size. The strategy distribution of NSPs and access networks strongly correlates with geographic expanse; the fraction of networks that announce an Open peering policy decreases with geographic expanse.

Strategy transitions: Between August 2010 and August 2013, 130 ASes in PeeringDB changed their peering strategy. Surprisingly, given the prevalence of Open peering, *most (70%) of networks that changed their peering strategy moved towards a more selective strategy.* When classified by self-reported business type,

among these 130 ASes, 71% of access providers, 52% of content providers, and 80% of transit providers became more selective in their peering. When classified by traffic volume, 61% of networks with reported traffic less than 1Gbps, 81% of networks with 1-100Gbps and 83% of networks with more than 100Gbps became more selective in their peering policies. A plausible hypothesis for the change towards more selective peering is that these ASes became less profitable due to Open peering, causing them to revert back. This shift may also be evidence of the "peering life cycle" [27], where networks initially advertise an Open peering policy and then become more selective with time.

6. RELATED WORK

To the best extent of our knowledge this work is the first comprehensive look at PeeringDB data with the goal of extracting usable network-specific properties that could help Internet researchers parameterize and validate their models [5, 6, 7, 8]. Several schemes for classifying ASes according to business type were proposed [3, 2, 4]; they all relied on hand-classifying a few networks and using these labels as ground truth to support inference of business types of other networks, an area in which PeeringDB data can be very useful. In the area of measuring and characterizing IXPs, Augustin et al. [28] developed techniques based on targeted traceroutes to map connectivity at IXPs. Ager et al. [29] used traffic data from a large European IXP to measure and characterize peering connections. Silvius [30] did a comparative survey of IXPs, focusing on the differences in the IXP ecosystem between North America and Europe. Our work complements these studies by providing a *network-centric* view as seen through PeeringDB. Norton surveyed a set of 28 peering policies [31] to identify differences in peering behavior between networks of different types and sizes. Norton also discusses how the peering preferences of networks change over time [27, 32]. However, that work was based on small-scale surveys of network operators. PeeringDB gives us the opportunity to study similar questions at a much larger scale.

7. CONCLUSIONS AND FUTURE WORK

Our goal was to explore PeeringDB, an online reource widely used by operators for peering, and assess its applicability to Internet research. We found that the network membership in PeeringDB is representative of the general population of transit, content, and access providers, especially the more prominent networks of each type. While PeeringDB is useful for network-specific properties, its use for measuring IXP-specific properties is questionable, as many networks in developing regions do not register in PeeringDB. From PeeringDB snapshots we extracted information about peering policies, geographic expanse, traffic volume, and correlations among these properties. We envision that this data will help to better parameterize and validate models of interdomain interconnection and dynamics, our own work [6, 33] as well as that of other researchers. More work needs to be done to validate self-reported network properties in PeeringDB. In that direction, we are focusing on comparing network-advertised peering policies with their peering behavior seen at IXP route servers.

8. REFERENCES

[1] "PeeringDB," http://www.peeringdb.com.
[2] A. Dhamdhere and C. Dovrolis, "Twelve Years in the Evolution of the Internet Ecosystem," *IEEE/ACM Transactions on Networking*, 2011.
[3] X. Dimitropoulos, D. Krioukov, G. Riley, and k. claffy, "Revealing the Autonomous System Taxonomy: The Machine Learning Approach," in *Proc. PAM*, Mar 2006.

[4] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (in)completeness of the observed internet as-level structure," *IEEE/ACM Trans. on Networking*, Feb. 2010.

[5] H. Chang, S. Jamin, and W. Willinger, "To Peer or Not to Peer: Modeling the Evolution of the Internets AS-level Topology," in *Proc. IEEE INFOCOM*, 2006.

[6] A. Dhamdhere and C. Dovrolis, "The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh," in *Proc. ACM CoNEXT*, 2010.

[7] P. Holme, J. Karlin, and S. Forrest, "An Integrated Model of Traffic, Geography and Economy in the Internet," *ACM SIGCOMM CCR*, 2008.

[8] J. Corbo, S. Jain, M. Mitzenmacher, and D. C. Parkes, "An Economically-Principled Generative Model of AS Graph Connectivity," in *Proc. IEEE INFOCOM*, 2009.

[9] "University of Oregon Route Views Project," http://www.routeviews.org/.

[10] RIPE NCC, "Routing Information Service (RIS)," http://www.ripe.net/ris/.

[11] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and K. Claffy, "AS Relationships, Customer Cones, and Validation," in *Proc. ACM SIGCOMM IMC*, 2013.

[12] Cooperative Association of Internet Data Analysis, "CAIDA Data," http://www.caida.org/data.

[13] "List of Tier-1 ASes," http://en.wikipedia.org/wiki/Tier_1_network#List_of_tier_1_networks.

[14] "List of Tier-2 ASes," http://en.wikipedia.org/wiki/Tier_2_network.

[15] "Alexa: The web information company," http://www.alexa.com.

[16] "The Pirate Bay," http://thepiratebay.sx/.

[17] J. Snijders, "PeeringDB Accuracy: Is blind faith reasonable?" *NANOG 58*, 2013.

[18] Cisco Systems, "Cisco visual networking index: Forecast and methodology, 2012-2017."

[19] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet Inter-domain Traffic," in *Proceedings of ACM SIGCOMM*, 2010.

[20] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse?" in *Proc. PAM*, 2008.

[21] Google Inc., "Peering & content delivery," https://peering.google.com/about/peering_policy.html.

[22] Akamai Technologies, "Peering information," http://www.akamai.com/peering/.

[23] Limelight Networks, "Limelight networks peering guidelines," http://login.llnw.net/noauth/peering.cgi.

[24] V. Giotsas, S. Zhou, M. Luckie, and K. Claffy, "Inferring multi-lateral peering," in *Proceedings of ACM CoNEXT*, 2013.

[25] A. Lodhi, A. Dhamdhere, and C. Dovrolis, "Peering Strategy Adoption by Transit Providers in the Internet: A Game Theoretic Approach," in *Proceedings of W-PIN*, 2012.

[26] ——, "Open Peering by Internet Transit Providers: Peer Preference or Peer Pressure?" in *Proceedings of IEEE INFOCOM*, 2014.

[27] W. Norton, "The Lifecycle of a Peering Inclination," http://drpeering.net/white-papers/Peering-Policies/.

[28] B. Augustin, B. Krishnamurthy, and W. Willinger, "IXPs: mapped?" in *Proc. ACM SIGCOMM IMC*, 2009.

[29] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger, "Anatomy of a large European IXP," in *Proc. ACM SIGCOMM*, 2012.

[30] S. Silvius, "Internet Exchange Points: A closer look at the differences between continental Europe and the rest of the world," https://www.euro-ix.net/documents/894-ix.

[31] W. Norton, "A Study of 28 Peering Policies," http://drpeering.net/white-papers/Peering-Policies/.

[32] ——, *The Internet Peering Playbook: Connecting to the Core of the Internet.* DrPeering Press, 2012.

[33] A. Lodhi, A. Dhamdhere, and C. Dovrolis, "GENESIS: An Agent-Based Model of Interdomain Network Formation, Traffic Flow and Economics," in *IEEE INFOCOM*, 2012.