Quantifying the Completeness of the Observed Internet AS-level Structure

Ricardo Oliveira
UCLA
rveloso@cs.ucla.edu

Dan Pei
AT&T Labs Research
peidan@research.att.com

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

Beichuan Zhang
University of Arizona
bzhang@arizona.edu

Lixia Zhang
UCLA
lixia@cs.ucla.edu

ABSTRACT

Despite significant efforts to obtain an accurate picture of structure at the level of individual autonomous systems (ASes), much remains unknown in terms of the quality of the inferred AS maps that have been widely used by the research community. Building upon our recent results reported in [16], in this paper we take a first step towards quantifying the (in)completeness of the observed AS-level connectivity as seen by the commonly-used vantage points of RouteViews and RIPE-RIS. Calling the current set of vantage points the "public view," we developed a new heuristic to identify all the ASes whose AS-level connectivity is completely captured by the public view. Our results indicate that the public view is capable of revealing the full connectivity of only 4% of all the ASes, which accounts for 77% of all large ISPs and 34% of small ISPs, but only 0.5% of stub ASes. For the remaining 96% ASes, the public view captures their customer-provider links, but may miss most of their peer links. We also provide evidences that the bulk of the missing connectivity involves peer links below the line of sight of the public view, typically between stub ASes and small ISPs as well as among stub ASes. Our findings call for new ways of inferring AS-level connectivity that do not rely solely on the use of active/passive measurements from vantage points, and our preliminary results towards this direction look promising.

1. INTRODUCTION

There exists a number of research efforts (see for example [10, 12, 9, 20, 18, 19]) that aim to identify, quantify, and understand the inherent limitations of the Internet AS maps inferred from publicly available datasets provided by Route Views and RIPE-RIS. As a new contribution to the existing efforts, we conducted a number of case studies that yielded new insights into which parts of the actual AS topology are adequately captured in these maps and which parts are missing and why, as reported in a recent paper [16]. Calling the commonly used vantage points of Route Views and RIPE-

RIS the "public view," we showed that this public view (i) accounts for the full connectivity of all the Tier-1 ASes, (ii) captures all customer-provider links in the Internet, provided that one includes the historical data from the public view, and (iii) misses a large number of peer links, especially in the lower tiers of the Internet routing hierarchy.

Building upon the above findings, in this paper we take a first step towards quantifying how much of the AS-level topological connectivity may be missing from the public view. Based on the no-valley routing policy, we first develop a new heuristic to accurately identify all the customer-provider links in the observed AS topology; we then classify ASes into customer-provider relations. Since a customer AS can observe all the AS links of its provider(s) over a long enough time period, by identifying those ASes who have at least one customer AS that hosts a vantage point in the public view, we are able to identify all the ASes whose connectivity is captured in the public view.

Our results show that the public view is capable of revealing the full peer connectivity of only 4% of all the ASes, which accounts for 77% of all the large ISPs and 34% of the small ISPs, but only 0.5% of the stub ASes. For the remaining 96% ASes, the public view captures their customer-provider links, but may miss their peer links. Although it is generally believed that the public view misses a large portion of AS links in the Internet topology, we believe that we are the first to be able to *quantify* exactly how much is missing.

Our findings are not very encouraging for settling the (in)completeness problem of inferred AS-level topology maps. For one, the part of the actual AS-level topology that are largely uncovered by the public view is the edge ASes' connectivity, and the large number of edge ASes makes it infeasible to install vantage points in all of them, not to mention a number of potential non-techical issues involved in doing so. To make things worse, the edge connectivity is precisely the part of the AS topology that changes most[17]. One driving factor is the aggressive peering among ASes to reduce cost and improve performance. The ever increasing connectiv-

ity density around the edges plays a key role in the Internet topology evolution. Not only is the incomplete part hard to capture, it also represents a moving target.

Faced with this dilemma when it comes to obtain complete Inferred AS maps, our results call for alternative approaches to inferring AS-level connectivity that do not solely rely on the use of vantage points (through either active or passive measurements), but is based more on the fact that ASes are involved in business relationship, and infers AS connectivities from knowing their business model, economic health, geographic extent, carried traffic, etc.. Compared to the largely AS-agnostic inference approaches that have been considered so far, more AS-aware methods have the promise to capture the key forces at work more accurately in the actual AS-level eco-system. More specifically, in contrast to the work of Ratz *et al.* [18], we are able to provide certain absolute bounds on the observed AS level connectivity based on AS classification and link classification, and to pinpoint where in the hierarchy these links are missing. Such methods open up new opportunities for measurement, inference, and modeling.

2. DATA SETS

In this paper we use two types of data to infer AS relationships, classify ASes, and infer ASes' presence at IXPs.

BGP data: Throughout this paper, we mainly use BGP data from Routeviews[8] and RIPE-RIS[7] collected over a 7-month period from 2007-06-01 to 2007-12-31. We term this data set *public view*. This data include BGP tables and updates from ∼700 operational routers in ∼400 ASes, although only about 100+ routers from each of the two sources have full BGP routing table data as indicated by Figure 1. Due to the overlap between Routeviews and RIPE-RIS in their monitored ASes, and due to the fact that some ASes have multiple monitors, the set of routers with full tables correspond to only 126 unique ASes. All Tier-1 ASes are included in this set except AS209 (Qwest); fortunately one of AS209's customers has a monitor which can observe all AS209's connectivity over time. Although there are additional BGP data sources such as route servers, looking glasses and the Internet Routing Registry [3], the amount of additional AS connectivity they uncover is incremental, so we do not use them here. Furthermore, these extra data sources often only provide partial BGP tables (and no updates), and, as we plain in Section 5, our heuristic for an accurate quantification of the observed completeness requires vantage points with full routing tables. Note that we currently do not use AS topological data derived from traceroute measurements due to the well known issues in converting router paths to AS paths, which have been extensively reported in previous work [11, 15, 13, 17].

IXP data: There are a number of websites, such as Packet Clearing House (PCH) [4], Peeringdb [5], and Euro-IX [1], that maintain a list of IXPs worldwide and also provide a list of ISP participants in some of the IXPs. The list of IXP fa-
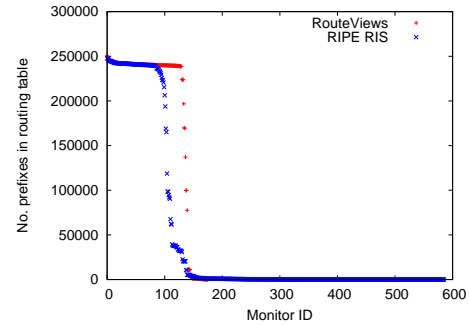

Figure 1: Table size of monitors in public view.

Presences (AS-IXP pairs)	Peeringdb	Euro-IX	PCH
Listed on source website	2,552	2,203	862
Inferred from reverse DNS	3,043		3,618
Unique within the source	4,442	2,203	3,968
Total unique across all sources	6,215		

Table 1: IXP membership data, December 2007.

cilities is believed very close to being complete [6], the list of ISP participants at the IXPs may be either incomplete or outdated since it is provided by the ISPs on a voluntary basis. However, (1) most IXPs publish the subnets they use in their layer-2 clouds, and (2) the best current practice [2] recommends that each IXP participant keeps reverse DNS entries for their assigned IP addresses inside the IXP subnet and no entries for unassigned addresses. Based on the combination of (1) and (2), we adopted the method used in [19] to infer IXP participants. The basic idea is to do reverse DNS lookups on the IXP subnet IPs, and then infer the participant ISPs from the returned domain names.

We define an (AS, IXP) pair as a *presence*. For example, if both AS A and AS B peer at IXP X, there are two presences: (A,X) and (B,X). From the aforementioned three data sources, we were able to derive a total of 6,215 unique presences corresponding to 2,843 ASes in 177 IXPs worldwide. Table 1 shows the breakdown of the observed presences. From the list of presences inferred from DNS for Peeringdb, 491 were already in its participant list, and from the presences inferred from DNS for PCH, 512 were already in the participant list.

3. NETWORK CLASSIFICATION

In this section we describe a novel method to infer the business relationships between ASes, and our process of using this method to classify different ASes into classes.

3.1 Inferring AS Relationships

Our recent work [16] reported that monitors at the top of the routing hierarchy (*i.e.* Tier-1 monitors) are able to reveal all the downstream provider-customer connectivity over time [1]. This is an important observation because, by definition, each non-Tier-1 AS is a customer of (or downstream

---

[1]Assuming routes follow a no-valley policy.

of) at least one Tier-1 AS, and essentially all the provider-customer links in the Internet can be observed at the Tier-1 monitors over time. This is the basic idea of our AS relationship inference algorithm.

Our algorithm assumes the set of Tier-1 ASes are already known[2]. By definition of Tier-1 ASes, all the links between Tier-1 ASes are peer links, and a Tier-1 AS is not a customer of any other ASes. Suppose a monitor at Tier-1 AS m reveals an ASPATH m-a_1-a_2-...-a_n. The link m-a_1 can be either a provider-customer link, or a peer link since sometimes a Tier-1 might have a specially arranged peer relationship with a lower-tiered AS. However, according to the no-valley policy, a_1-a_2, a_2-a_3, ... , a_{n-1}-a_n must be provider-customer links, because a peer or provider route should never be propagated upstream from a_1 to m, therefore the segment a_2, ..., a_n must correspond to a customer route received by a_1. How can one infer the relationship of m-a_1 link? According to the no-valley policy, if m-a_1 is provider-customer link, this link should appear in the routes propagated from m to other Tier-1 ASes, and whose monitors will show this link. On the other hand, if m-a_1 is a peer link, it should never be propagated to or seen by monitors at other Tier-1 ASes (other than m itself). Given we have monitors in all but one Tier-1 ASes, we can accurately infer the relationship m-a_1 by looking at whether it is revealed by other Tier-1 ASes besides m. Using the above method, we can find and label all the provider-customer links, while all other links revealed by all monitors are then labeled as peer links.

Our algorithm is illustrated in Figure 2, where 1, 2, 3, and 4 are known to be Tier-1 ASes. Suppose Tier-1 AS 2 monitor reveals an ASPATH 2-5-6-8 and another ASPATH 2-7-9; while monitors at Tier-1 AS 4 reveals an ASAPTH 4-2-7-9, but none of 1, 3, 4 reveals an ASPATH ending at 2-5-6-8. According to the above algorithm, 5-6, 6-8, and 7-9 are definitely provider-customer links. 2-7 is provider-customer link since it is revealed by Tier-1 ASes other than 2, while 2-5 is peer link since it is not revealed by any other Tier-1 ASes. Furthermore, suppose AS 6 is a monitor and it reveals link 6-7, and 6-7 is never revealed by Tier-1s 1,2,3, or 4. Then this 6-7 is a peer link according to our algorithm.

From our measurements of Tier-1 routes over the 7-month period, we were able to infer a total of 70,698 provider-customer links. We also noticed some of these links were in routes that had a very short lifetime (less than 2 days). These cases are most likely caused by BGP misconfigurations (*e.g.* route leakages) or prefix hijacks, as described in [14]. After filtering out all the routes with a lifetime shorter than 2 days, we excluded 5,239 links, and ended up with a total of 65,459 provider-customer links.

3.2 AS classification

In this section we make use of the inferred provider-customer relations to classify ASes into several functional types. In

---

[2]The list of Tier-1 ASes can be obtained from website such as `http://en.wikipedia.org/wiki/Tier_1_carrier`
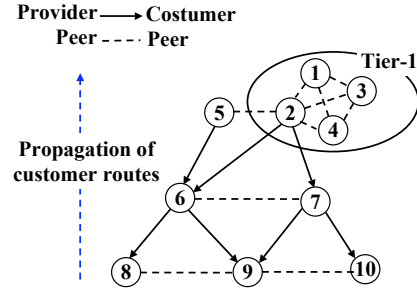

Figure 2: Provider-customer links are revealed over time to Tier-1 monitors.

the past this classification has been done based on the degree of an AS, or the number of prefixes originated by the AS. Unfortunately, the degree can be misleading since it includes a mix of providers, peers and customers in the count; the number of prefixes originated by an AS may also be not very meaningful since the length of the prefixes can be very different and the routes carried downstream may not be accounted.

To overcome these limitations, we use the number of downstream customer ASes (or "customer cone") to classify ASes; the number of downstream customers were extracted from the routes gleaned over the 7-month period from the Tier-1 monitors. Figure 3 shows the distribution of the number of downstream customers per AS. We note that about 80% of the ASes do not have any customers, and a significant fraction of ASes only have a very small number of customers. We thus label as *stub* those ASes with 4 or less customers, which encompass about 92% of all the ASes. This stub class should correspond to end networks which either do not provide transit at all, or offer very limited transit service to few local customers, *e.g.* universities providing transit connectivity for small research facilities. Further, based on the knee of the distribution in Figure 3, we label as *small ISPs* those ASes which have between 5 and 50 downstream customers. They correspond to about 6% of the total ASes. The remaining ASes in the long tail which are not known as Tier-1s are labeled as *large ISPs*. Table 3 shows the number of ASes in each class. We analyzed the sensitivity of the classification thresholds by changing the values slightly, which did not lead to significant difference in the end result.

4. COVERAGE OF THE PUBLIC VIEW

In this section we quantify the completeness of the AS topology as observed by the public view by a few measures. First we would like to note that according to observations made in [16], a monitor can discover all the connections of all the upstream ASes over time. For example, in Figure 2, a monitor at AS 7 will receive routes from upstream providers that will contain the peer links existing upstream, in this case the links 2-1, 2-3, 2-4 and 2-5 (in addition to the provider-customer links existing upstream). Therefore, by starting at
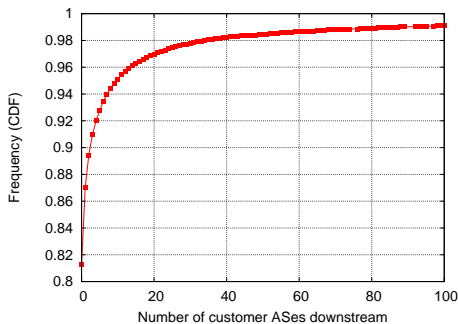
3

Figure 3: Distribution of number of downstream customers per AS.

AS 7 and following all provider-customer links upstream, we pass through the ASes that are *covered* by AS 7, in the sense that AS 7 is able to reveal all their connectivity. In Figure 2, the ASes covered by AS 7 is just AS 2, but AS 6 covers both AS 5 and AS 2. We have the following definition:

Covered AS: is an AS that can be reached by a monitor using a provider chain, e.g. a sequence of customer $\rightarrow$ provider links. If an AS X is covered, than all its connectivity, including the peer links, are revealed in the public view by the covering monitor(s). If the monitor is a router residing in X, then X is also considered covered.

We measured the number of ASes covered by the monitors in the public view, the results are shown in Table 2. For comparison purposes, we included both the set of monitors with full tables and full+partial tables, though the end result is very similar. The most striking observation is that the current set of monitors in public view is only able to cover 4% of the total number of ASes, which indicates that the widely used view used by the research community may in fact miss most of the *peer* connectivity in the network happening within the remaining 96% of the ASes.

We extend this analysis to prefixes and traffic volume in the following way. Assume AS i originates P_i prefixes, then we add up the prefixes for all ASes covered by AS i to produce the total number of prefixes covered by AS i: $\sum_{i \in cov} P_i$. That is, with our monitor set, we are able to cover the AS links used in all the routes to these prefixes. According to Table 2, at least 22% of the prefixes are reachable through AS links already covered. This number should be taken as a lower bound, since there can be prefixes not covered that are reached always through already covered routes.

In order to extend this analysis to traffic, we make use of proprietary Netflow data from a Tier-1 backbone. We denote by f_i the total fraction of traffic received by the P_i prefixes originated by AS i (which we can easily extract from the Netflow data)[3]. Then the fraction of covered traffic is

---

[3]We can view f_i as a measure of popularity of AS i.

Parameter	Full tables	Full+partial tables
No. monitored ASes	121	411
No. ASes	1,101 / 28,486 $\simeq$ 4%	1,552 / 28,486 $\simeq$ 5 %
Prefixes	52,861 / 236,237 $\simeq$ 22%	60,987 / 236,237 $\simeq$ 26%
Traffic	$\simeq$ 22%	$\simeq$ 25%

Table 2: Coverage of BGP monitors.

Type	ASes	Monitored ASes	Covered ASes	
			aggregated	by covering type
Tier-1	9	8	9 (100%)	8
Large ISP	436	45	337 (77.3%)	954
Small ISP	1,829	36	629 (34.4%)	269
Stubs	26,209	37	126 (0.5%)	160

Table 3: Coverage of BGP monitors for different network types.

given by $\sum_{i \in cov} f_i$, which is about 22% according to Table 2. Again, this should be viewed as a lower bound, in the sense that at least 22% of the traffic in the network follow AS links already covered by the monitor set.

Finally, we do an analysis of the covered ASes in terms of their classes, which is shown in Table 3. The column "Covered ASes-aggregated" refers to the total fraction of covered ASes in each class, whereas the column "Covered ASes-by covering type" refers to the number of ASes covered by the monitors in each class. For instance, 77.3% of the large ISPs are covered by monitors, and monitors in large ISPs cover 954 total ASes. The numbers in the table indicate that Tier-1s are fully covered, large ISPs are mostly covered, small ISPs remain largely uncovered (just 34.4%), and stubs are almost completely uncovered (99.5%). This is because most of the monitors reside in the core of the network, and in order to cover a stub, we would need to place a monitor in the stub or in any of its downstreams which is unfeasible to do at the scale of the Internet due to the very large number of stubs in the network.

The public view captures all connectivity of a covered AS, including all *of its peer links. For an AS not being covered, the public view captures all of its customer-provider links and some peer links, but may miss most of their peer links.* For example, if AS C is covered but AS U and V are not, then a peer link C–U will be captured by the public view, but the peer link U–V will not. Therefore, how much connectivity the public view misses depends on how many *peer* links exist between the 96% ASes that are not covered. The more such peer links, the more connectivity the public view misses. Our definition of coverage provides an upper bound estimate for the missing connectivity by the public view.

5. QUANTIFYING THE INVISIBLE CONNECTIVITY

In the previous section we analyzed how complete was the view of the current set of public monitors in terms of covered ASes, prefixes and traffic. In this section we look into estab-
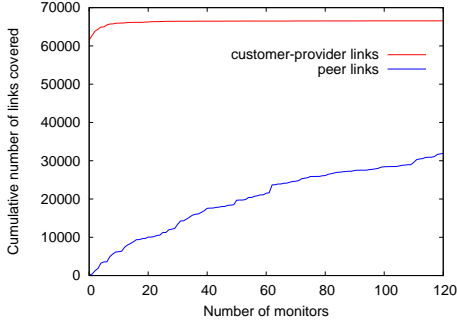
Figure 4: Links captured using monitors with full BGP tables, 7-month period.

Figure 5: Modeling the exposure of peer links.

lishing the bound on the number of peer links that are missing from observation. Figure 4 shows the cumulative number of unique customer-provider and peer links captured by the monitors with full tables when picked by random ordering. We can clearly observe that customer-provider links are all covered after a few monitors[4]. However, when we look at the curve for peer links, we notice a steady increase as we add more monitors, in the sense that each monitor adds new peer links that were hidden before. The challenge now lies in estimating how much connectivity in terms of peer links remains invisible after adding the nth monitor. We estimate this value by using the following simple model. Assume there are N ASes in the network and each can potentially have a monitor to provide BGP routing tables and updates to public view. Now suppose we keep collecting peer links by looking at these monitors, one after another other in a random order, and we want to know how many peer links are revealed after looking at n monitors. The number of visible peer links after observing n monitors, $V(n)$ is given by:

$$V(n) = \sum_i p_i(n) = \sum_i (1 - \frac{\binom{N-S_i}{n}}{\binom{N}{n}}) \qquad (1)$$

where $p_i(n)$ is the probability to cover a specific link i after observing n monitors. $p_i(n)$ is given by the hypergeometric distribution, where S_i is the number of monitors that use link i and $\frac{\binom{N-S_i}{n}}{\binom{N}{n}}$ is the probability that link i remains hidden after looking at n monitors.

Figures 5 reproduces the curve of peer links of Figure 4. We note that after a certain number of monitors, the number of collected peer links follows a linear increase trend that we can explain by our simple model. Assuming $n \ll N$, the hypergeometric distro in Equation 1 can be replaced by a binomial distro having parameter $s_i = \frac{S_i}{N}$. We term s_i the *scope* of link i. Therefore, we would have:

$$V(n) \simeq \sum_i (1 - (1-s_i)^n) \simeq L - \sum_i (1-s_i)^n \qquad (2)$$

---

[4]If the first monitors were Tier1s, they would be covered immediately, as we saw in previous section.
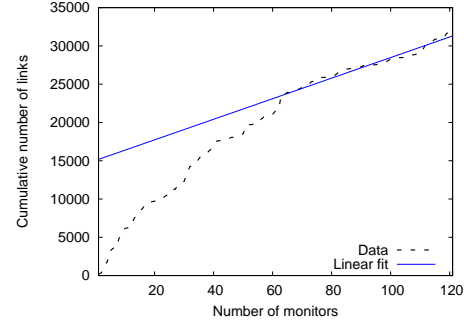
where L is the total number of peer links in the network.

We have been assuming that all nodes have the same chance of being randomly picked as a monitor, where in reality just by looking at Table 3, we note that there is a bias to pick monitors that are at the core of the network. Therefore, to account for this bias, we sort monitors into transits and stubs, and treat these cases separately. Given the current monitor set has $w_t = 89$ transits (Tier1+small+large ISPs) and $w_s = 37$ stubs, we assume a new monitor will be a stub with chances $q = \frac{w_s N_s}{w_s N_s + w_t N_t}$ and a transit with chances $1 - q$, where N_s is the number of remaining stubs in the network and N_t the number of remaining transits. Therefore, we can rewrite $s_i = q s_{i,s} + (1-q) s_{i,t}$, where $s_{i,s}$ is the fraction of stubs that use link i, and $s_{i,t}$ are the fraction of transits that use link i.

Lets denote $H(n) = \sum_i (1-s_i)^n$ the number of hidden links after observing n nodes. Note that for a sufficiently large number of monitors n, the number of hidden links $H(n)$ can be well approximated by considering only the contribution of very small scope links since the links with high scope were probably already revealed. Lets assume these hidden links have a very small probability $s_{i,t} = s_{i,s} = \beta$ of being revealed by a given monitor, then we can write:

$$H(n) \simeq A(1-\beta)^n \simeq A(1 - n\beta)$$

where A is the total number of links with very small scope. Therefore, we can rewrite Equation 2 as:

$$V(n) \simeq L - A(1 - n\beta) \simeq (L - A) + A\beta n$$

which explains the linear trend of the curve in Figure 5. The parameters $(L-A)$ and $A\beta$ can be estimated by curve fitting the measured data in Figure 5. The parameter β quantifies the chances of a monitor to reveal a hidden peer link (after looking into a large number of monitors). In the worst case, for a peer link between two stubs, the link is only revealed by the two incident monitors: $\beta = q\frac{2}{N_s}$. On the other hand, in the best case $\beta \simeq 1/n$, where n is the number of monitors available, *i.e.* the hidden peer link is at least as hard to capture as the number of monitors used so far to capture it. So we have $q\frac{2}{N_s} < \beta < \frac{1}{n}$. In Figure 6 we show a projection of the number of total peer links L depending on the value of
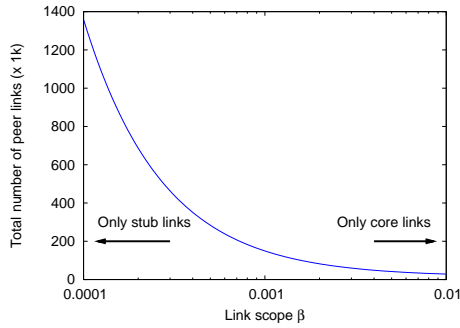
5

Figure 6: Establishing bounds for the number of peer links based on the link scope parameter β.

Figure 7: Projection of the number of peer ASes of a content provider and comparison with ground truth.

β. If we are on the left side of the curve, then all remaining hidden peer links are at the edge of network between stub ASes. If we are on the right side, then all peer links are in the core and already captured by our monitors. Comparing these values with the $\sim 30k$ peer links captured by the current monitors, we estimate that the monitors's view might be missing roughly up to 90% or more of the total peer connectivity.

6. LARGE CONTENT NETWORKS

Since majority of the peer links are missing from public view and it is impossible to install a BGP monitor in every AS in the network, new methods are required to fill in the missing peer links to achieve a complete and accurate AS level map. In this section we develop a method towards this goal. We focused on large content networks in this paper because previous work [16] has shown their peer links are mostly missing from public view thus we are focusing on a challenging special case. Another reason is that we happen to have access to the ground truth for one large content provider C so that we can evaluate our heuristics against the ground truth at C.

Peering can be implemented in two ways: *private peering* and *public peering*. A private peering is a dedicated router-to-router layer-2 connection between two networks. Private peering provides dedicated bandwidth, is easier to troubleshoot problems, but has higher cost. Recently there is a trend to migrate private peerings to public peerings since the latter costs less and its bandwidth capacity is increasing. Public peering usually happens at the Internet Exchange Points (IXPs), which are third-party maintained physical infrastructures that enable physical connectivity between their member networks. Currently most IXPs connect their members through a common layer-2 switching fabric (or layer-2 cloud). Though IXPs enable physical connectivity between all participants, whether to establish BGP peering sessions on top of the physical connectivity is up to individual networks. It is possible that one network may only peer with some of the other participants in the same IXP.
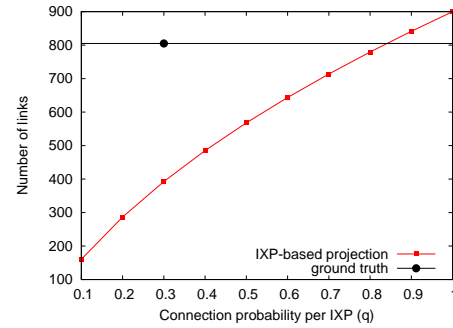
Large content networks are a special case of networks that usually engage in heavy public peering at IXPs. These networks usually have a small number of downstream customers and a small incoming/outgoing traffic ratio. Since their main business is not to provide transit but rather to enable access to their content, these networks usually have a very open peering policy, peering with whoever wants to peer with them. These policy has two benefits, first by having direct connection with peers they speed up the content delivery and second they save on traffic sent upstream (reducing their Internet access cost).

We now develop a method to infer the public peers of a given content provider C, for which we obtained ground truth information after conversations with its network operators, who also disclosed that C peers with 80-90% of the participants at each IXP. We assume that in each IXP where C has presence, it connects to a fixed fraction q of the networks also colocated at that IXP, *i.e.* if C has n common locations with another network X, then the chances that C and X are connected in at least one IXP are given by $1-(1-q)^n$. More generally, the expected number of peer ASes of C, P_C is given by $P_C = \sum_i (1 - (1 - q)^{n_i})$, where i represents all networks that have at least one common presence with C, and n_i is the number of IXPs where both C and i have presence. In our IXP data set, C has presence in 30 IXPs worldwide, which is very close to the number that was disclosed to us by the operators of C. Based on the IXP data and our above model, we plot the projected number of peer ASes for C in Figure 7, where we also show the ground truth. We note that our projection, given the $q = 80 - 90\%$ for C, is very close to the ground truth.

With this in mind, and given the open peering policy of the content networks, a feasible approach to fill the missing connectivity for large content providers would be to simply assume they are connected to all the participants that share at least one common IXP with them. If we follow this approach for the special case of C, we would end up with an accuracy of about 85%, *i.e.* about 15% of the peers would be false positives. We are currently investigating different approaches to fill the missing connectivity for other types of ASes such as small and large ISPs.

6

7. CONCLUSION

Although it is generally believed that the public view misses a large portion of AS links in the Internet topology, to the best of our knowledge this paper represents the first attempt to *quantify* exactly how much may be missing. Through the use of a new heuristic developed in this paper, we show that the public view is capable of capturing the full connectivity of only 4% of all the ASes. However this low percentage of covered ASes should be viewed together with the results reported in [16], i.e., the public view is capable of capturing all the customer-provider AS links, and most of peer links between large ISPs, in the topology over time. The bulk of the potentially missing connectivity involves peer links bellow the line of sight of the public view. We presented a simple model that provides the upper bound estimate on the number of potentially missing peer links. This is only the initial step towards filling the missing gaps in the AS maps widely used by the research community and should be viewed as a stimulus towards a more comprehensive approach, including new inference techniques that do not rely uniquely on information provided by a small set of vantage points, and new modeling that takes into consideration AS' economic incentives.

8. REFERENCES

[1] European Internet exchange association. http://www.euro-ix.net.
[2] Good practices in Internet exchange points. http://www.pch.net/resources/papers/ix-documentation-bcp/ix-documentation-bcp-v14en.pdf.
[3] Internet Routing Registry. http://www.irr.net/.
[4] Packet clearing house IXP directory. http://www.pch.net/ixpdir/Main.pl.
[5] PeeringDB website. http://www.peeringdb.com/.
[6] Personal Communication with Bill Woodcock@PCH.
[7] RIPE routing information service project. http://www.ripe.net/.
[8] RouteViews routing table archive. http://www.routeviews.org/.
[9] H. Chang. *An Economic-Based Empirical Approach to Modeling the Internet Inter-Domain Topology and Traffic Matrix*. PhD thesis, University of Michigan, 2006.
[10] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. Towards capturing representative AS-level Internet topologies. *Elsevier Computer Networks Journal*, 44(6):737–755, 2004.
[11] H. Chang, S. Jamin, and W. Willinger. Inferring AS-level Internet topology from router-level path traces. In *SPIE ITCom*, 2001.
[12] H. Chang and W. Willinger. Difficulties measuring the Internet's AS-level ecosystem. In *Annual Conference on Information Sciences and Systems (CISS'06)*, pages 1479–1483, 2006.
[13] Y. Hyun, A. Broido, and kc claffy. On third-party addresses in traceroute paths. In *Proc. of Passive and Active Measurement Workshop (PAM)*, 2003.
[14] R. Mahajan, D. Wetherall, and T. Anderson. Understanding BGP Misconfiguration. In *In Proc. of ACM SIGCOMM*, 2002.
[15] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz. Towards an accurate AS-level traceroute tool. In *Proc. of ACM SIGCOMM*, 2003.
[16] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. In Search of the elusive Ground Truth: The Internet's AS-level Connectivity Structure. In *Proc. ACM Sigmetrics*, 2008.
[17] R. Oliveira, B. Zhang, and L. Zhang. Observing the evolution of Internet AS topology. In *ACM SIGCOMM*, 2007.
[18] D. Raz and R. Cohen. The Internet dark matter: on the missing links in the AS connectivity map. In *Proc. of IEEE INFOCOM*, 2006.
[19] Y. He, G. Siganos, M. Faloutsos, S. V. Krishnamurthy. A systematic framework for unearthing the missing links: measurements and impact. In *Proc. of NSDI*, 2007.
[20] B. Zhang, R. Liu, D. Massey, and L. Zhang. Collecting the Internet AS-level topology. *ACM SIGCOMM Computer Comm. Review (CCR)*, 35(1):53–61, 2005.