The (In)Completeness of the Observed Internet AS-level Structure

Ricardo Oliveira, *Member, IEEE*, Dan Pei, Walter Willinger, *Fellow, IEEE*, Beichuan Zhang, and Lixia Zhang, *Fellow, IEEE*

***Abstract*—Despite significant efforts to obtain an accurate picture of the Internet's connectivity structure at the level of individual autonomous systems (ASes), much has remained unknown in terms of the quality of the inferred AS maps that have been widely used by the research community. In this paper, we assess the quality of the inferred Internet maps through case studies of a sample set of ASes. These case studies allow us to establish the ground truth of connectivity between this set of ASes and their directly connected neighbors. A direct comparison between the ground truth and inferred topology maps yield insights into questions such as which parts of the actual topology are adequately captured by the inferred maps, which parts are missing and why, and what is the percentage of missing links in these parts. This information is critical in assessing, for each class of real-world networking problems, whether the use of currently inferred AS maps or proposed AS topology models is, or is not, appropriate. More importantly, our newly gained insights also point to new directions towards building realistic and economically viable Internet topology maps.**

***Index Terms*—Border Gateway Protocol (BGP) , interdomain routing, Internet topology.**

I. Introduction

MANY research projects have used a graphic representation of the Internet topology, where nodes represent autonomous systems (ASes) and two nodes are connected if and only if the two ASes are engaged in a business relationship to exchange data traffic. Due to the Internet's decentralized architecture, however, this AS-level construct is not readily available and obtaining accurate AS maps has remained an active area of research. All the AS maps that have been used by the research community have been inferred from either Border Gateway Protocol (BGP)-based or traceroute-based data. Unfortunately, both types of measurements are more a reflection of what we can measure than what we want to measure, as both have fundamental limitations in their ability to reveal the Internet's true AS-level connectivity structure.

While these limitations inherent in the available data have long been recognized, there has been little effort in assessing the degree of completeness or accuracy of the resulting AS maps. Although it is relatively easy to collect a more or less complete set of ASes, it has proven difficult, if not impossible, to collect the complete set of inter-AS links. The sheer scale of the Internet makes it infeasible to either install monitors everywhere or crawl the topology exhaustively. At the same time, big stakeholders of the AS-level Internet, such as Internet service providers and large content providers, tend to view their AS connectivity as proprietary information and are in general unwilling to disclose it. As a result, the quality of the currently used AS maps has remained by and large unknown. Yet numerous projects have been conducted using these maps of unknown quality, causing serious scientific and practical concerns in terms of the validity of the claims made and accuracy of the results reported.

In this paper, we take a first step towards a rigorous assessment of the quality of the Internet's AS-level connectivity maps inferred from public BGP data. Realizing the futility of attempting to obtain the complete global AS-level topology, we take an indirect approach to address the problem. Using a small number of different types of ASes whose complete AS connectivity information can be obtained, we conduct case studies to compare their actual connectivity with that of what we call the "public view" – the connectivity structure inferred from all the publicly available and commonly used BGP data source (i.e., routing tables, updates, looking glasses, and routing registry). These case studies enable us to understand and verify what kinds of AS links are adequately captured by the public view and what kinds of (and how many) AS links are missing from the public view. They also provide new insights into where the missing links are located within the overall AS topology.

More specifically, this paper makes the following original contributions. After we define what we mean by "ground truth" of AS-level Internet connectivity between a single AS and its neighbors in Section II, we report in Section III on a series of case studies which highlight the difficulties in establishing the ground truth, namely the data sources necessary to establish the AS-level connectivity are not publicly available for most ASes. Nevertheless, by classifying ASes into a few major types, we can explore what types and what fraction of each type of AS connectivity are missing from currently used AS maps, and we can typically identify the reasons why they are missing.

The main findings of our search for the elusive ground truth of AS-level Internet connectivity can be summarized as follows. First, inferred AS maps based on single *snapshots* of publicly available BGP-based data are typically of low quality. The percentage of missing links can range from 10%–20% for Tier-1 and Tier-2 ASes to 85% or more for large content networks. Second, the quality of the inferred AS maps can be significantly improved by including historic data of BGP updates from all

Manuscript received September 03, 2008; revised January 13, 2009 and March 21, 2009; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor O. Bonaventure. First published August 11, 2009; current version published February 18, 2010.

R. Oliveira and L. Zhang are with the University of California, Los Angeles, CA 90024 USA (e-mail: rveloso@cs.ucla.edu; lixia@cs.ucla.edu).

D. Pei and W. Willinger are with AT&T Labs–Research, Florham Park, NJ 07932 USA (e-mail: peidan@research.att.com; walter@research.att.com).

B. Zhang is with the University of Arizona, Tucson, AZ 85721 USA (e-mail: bzhang@arizona.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2009.2020798

existing sources. For example, links on backup paths can be revealed by routing dynamics over time, but the time period required to collect the necessary information can be several years. Third, through the use of data collected over long enough time periods, the public view captures all the links of Tier-1 ASes and almost all the customer-provider links at all tiers in the Internet. Fourth, due to the *no-valley* routing policy and the lack of monitors in most stub networks, the public view misses a great number of peer links at all tiers except Tier-1. It may miss up to 90% of peer links in the case of large content provider networks, which have been aggressively adding peer links in recent years.

The paper concludes with a discussion in Section V on several main lessons learned from our case studies, a brief review of related work in Section VI, and a summary detailing our future research plans in Section VII.

II. Searching for the Ground Truth

This section gives a brief background on interdomain network connectivity, defines its *ground truth*, describes the various data sets and methods that we used to infer the interdomain connectivity.

A. Interdomain Connectivity and Peering

As of summer 2008, the Internet consists of more than 27 000 networks called "Autonomous Systems" (AS). Each AS is represented by a unique numeric AS number and may advertise one or more IP address prefixes. ASes run the BGP [35] to propagate prefix reachability information among themselves. In the rest of the paper, we call the connection between two ASes an *AS link* or simply a *link*. As a path-vector protocol, BGP includes in its routing updates the entire AS-level path to each prefix, which can be used to infer the AS-level connectivity. Projects such as RouteViews [12] and RIPE-RIS [11] host multiple data *collectors* that establish BGP sessions with operational routers, which we term *monitors*, in hundreds of ASes to obtain their BGP forwarding tables and routing updates over time. BGP routing decisions are largely based on routing polices, in which the most important factor is the business relationship between neighboring ASes. Though the relationship can be fine-grained, in general there are three major types: *customer-provider*, *peer-peer* and *sibling-sibling*. In a customer-provider relationship, the customer pays the provider for transiting traffic from and to the rest of the Internet, thus the provider usually announces all the routes to the customer. In a peer-peer relationship, which is commonly described as "settlement-free," the two ASes exchange traffic without paying each other. However only the traffic originated from and destined to the two peering ASes or their downstream customers is allowed on a peer-peer link; traffic from their providers or other peers are not allowed. Therefore an AS does not announce routes containing peer-peer links to its providers or other peers. When an AS receives path announcements to the same destination from multiple neighbors, in general the AS prefers the path announced by a customer over that from a peer, and prefers a path from a peer over that from a provider. This is referred to as the **no-valley-and-prefer-customer** policy [22], which is believed to be a common practice in today's Internet. The sibling-sibling relationship is between two ASes that belong to the same organization, and is relatively rare; thus we do not consider it in this paper.

Among all the ASes, less than 10% are transit networks, and the rest are stub networks. A transit network is an Internet Service Provider (ISP) whose business is to provide packet forwarding service between other networks. Stub networks, on the other hand, do not forward packets for other networks. In the global routing hierarchy, stub networks are at the bottom or at the edge, and need transit networks as their providers to reach the rest of the Internet. Transit networks may have their own providers and peers, and are usually described as different tiers, e.g., regional ISPs, national ISPs, and global ISPs. At the top of this hierarchy are a dozen or so Tier-1 ISPs, which connect to each other in a fully mesh to form the core of the global routing infrastructure. The majority of stub networks today multihome with more than one provider, and some stub networks also peer with each other. In particular, *content networks*, e.g., networks supporting search engines, e-commerce, and social network sites, tend to peer with a large number of other networks.

Peering is a delicate but also important issue in interdomain connectivity. A network has incentives to peer with other networks to reduce the traffic sent to its providers, hence saving operational costs. But peering also comes with its own issues. For ISPs, besides additional equipment and management cost, they also do not want to establish peer-peer relationships with potential customers. Therefore ISPs in general are very selective in choosing their peers. Common criteria include number of colocations, ratio of inbound and outbound traffic, and certain requirements on prefix announcements [2], [1]. In recent years, with the fast growth of available content in the Internet, content networks have been keen on peering with other networks to bypass their providers. Because they have no concern regarding transit traffic or potential customers, content networks generally have an open peering policy and peer with a large number of other networks.

AS peering can be realized through either *private peering* or *public peering*. A private peering is a dedicated connection between two networks. It provides dedicated bandwidth, makes troubleshooting easier, but has a higher cost. 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[1]. Currently most IXPs connect their members through a shared layer-2 switching fabric (or layer-2 cloud). Fig. 1 shows an IXP that interconnects ASes A through G using a subnet 195.69.144.0/24. Though an IXP provides physical connectivity among all participants, it is up to individual networks to decide with whom to establish BGP sessions. It is often the case that one network only peers with some of the other participants in the same IXP. Public peering has a lower cost but its available bandwidth capacity between any two parties can be limited. However, with the recent increase in bandwidth capacity, we have seen a trend to migrate private peerings to public peerings.

B. Ground Truth Versus Observed Map

To study AS-level connectivity, we need a clear definition on what constitutes an inter-AS link. A link between two ASes exists *if the two ASes have a contractual agreement* to exchange traffic over one or multiple BGP sessions. The **ground truth**

[1]Note that private and public peering can happen in the same physical facility.

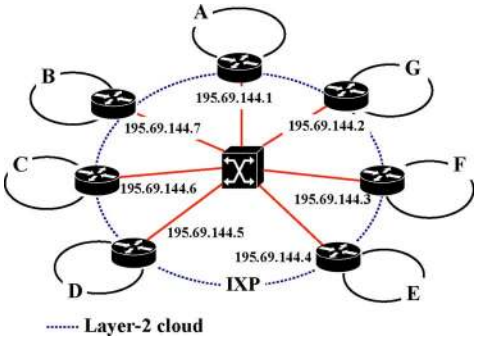


Fig. 1. A sample IXP. ASes A through G connect to each other through a layer-2 switch in subnet 195.69.144/24.

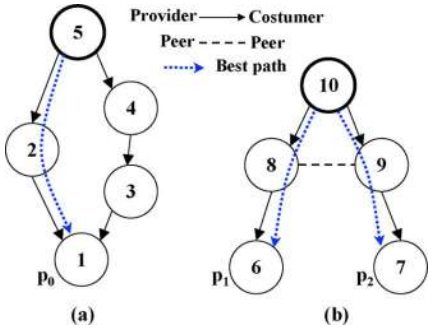


Fig. 2. A set of interconnected ASes, each node represent an AS. (a) Example of hidden links. (b) Example of invisible links.

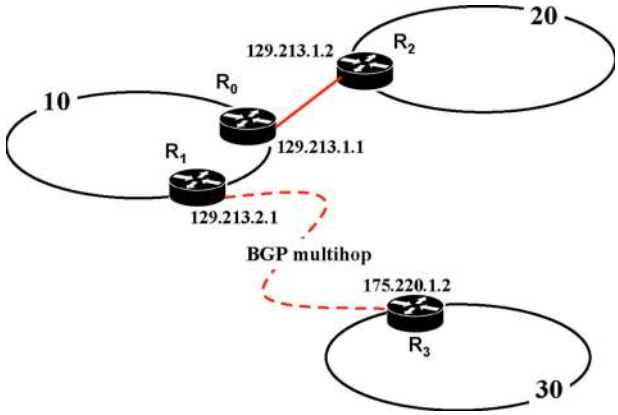


Fig. 3. Configuring remote BGP peerings. R_0 and R_2 are physically directly connected, while R_1 and R_3 are not.

of the Internet AS-level connectivity is the *complete* set of AS links. As the Internet evolves, its AS-level connectivity also changes over time. We use $G_{\text{real}}(t)$ to denote the ground truth of the entire Internet AS-level connectivity at time t.

Ideally if each ISP maintains an up-to-date list of its AS links and makes the list accessible, obtaining the ground truth would be trivial. However, such a list is proprietary and rarely available, especially for large ISPs with a large and changing set of links. In this paper, we derive the ground truth of several individual networks whose data is made available to us, including their router configurations, syslogs, BGP command outputs, as well as personal communications with the operators.

From router configurations, syslogs and BGP command outputs, we can infer whether there is a working BGP session, i.e., a BGP session that is in the *established* state as specified in RFC 4271 [35]. We assume there is a link between two ASes if there is at least one working BGP session between them. However if all the BGP sessions between two ASes are down at the moment of data collection, the link may not appear in the ground truth on that particular day, even though the two ASes have a valid agreement to exchange traffic. Fortunately we have continuous daily data going back for years, thus, the problem of missing links due to transient failures should be negligible. When inferring connectivity from router configurations, extra care is needed to remove stale BGP sessions, i.e., sessions that appear to be correctly configured in router configurations, but are actually no longer active. We use syslog data in this case to remove the stale entries (as described in detail in the next section). We believe that this careful filtering makes our inferred connectivity a very good approximation of the real ground-truth.

We denote an observed global AS topology at time t by $G_{\text{obsv}}(t)$, which typically provides only a partial view of the ground truth. There are two types of missing links when we compare G_{obsv} and G_{real}: **hidden links** and **invisible links**. Given a set of monitors, a hidden link is one that has not yet been observed but could possibly be revealed at a later time. An invisible link is one that is impossible to be observed by the given set of monitors. For example, in Fig. 2(a), assuming that AS5 hosts a monitor (either a BGP monitoring router or a traceroute probing host) which sends to the collector all the AS paths used by AS5. Between the two customer paths to reach prefix p_0, AS5 picks the best one, [5-2-1], so we are able to observe the existence of AS links 2-1 and 5-2. The three other links, 5-4, 4-3, and 3-1, are *hidden* at the time, but will be revealed when AS5 switches to path [5-4-3-1] if a failure along the primary path [5-2-1] occurs. In Fig. 2(b), the monitor AS10 uses paths [10-8-6] and [10-9-7] to reach prefixes p_1 and p_2, respectively. In this case, link 8-9 is *invisible* to the monitor in AS10, because it is a peer link that will not be announced to AS10 under any circumstances due to the no-valley policy.

Hidden links are typically revealed if we build AS maps using routing data (e.g., BGP updates) collected over an extended period. However, a new problem arises from this approach: the introduction of potentially stale links; that is, links that existed some time ago but are no longer present. A empirical solution for removing possible stale links has been developed in [33]. To discover all invisible links, we would need additional monitors at *most*, if not all, edge ASes where routing updates can contain the peering links as permitted by routing policy. The issues of hidden and invisible links are shared by both BGP logs and traceroute measurements.

C. Data Sets

We use the following data sources to infer the AS-level connectivity and the ground truth of individual ASes.

1) BGP Data: The **public view (PV)** of the AS-level connectivity is derived from all public BGP data at our disposal. These data include BGP forwarding tables and updates from $\sim$ 700 routers in $\sim$ 400 ASes provided by Routeviews, RIPE-RIS, Abilene [14], and the China Education and Research Network [3], BGP routing tables extracted from $\sim$ 80 route servers, and "show ip bgp sum" outputs from $\sim$ 150 looking glasses located

TABLE I
IXP MEMBERSHIP DATA, JULY 2007

Presences (AS-IXP pairs)	Peeringdb	Euro-IX	PCH
Listed on source website	2,203	2,478	575
Inferred from reverse DNS	2,878		3,613
Unique within the source	4,092	2,478	3,870
Total unique across all sources	6,084		

worldwide. In addition, we use "show ip bgp" outputs from Abilene and Geant [5] to infer their ground truth. Note that we currently do not use AS topological data derived from traceroute measurements due to issues in converting router paths to AS paths, as extensively reported in previous work [18], [29], [24], [33]. For results reported in Section IV, we use Routviews and RIPE-RIS data collected over a 7-mo period from 2007-06-01 to 2007-12-31. Due to the overlap in covered ASes between Routeviews and RIPE-RIS and the fact that some ASes have multiple monitors, the set of monitors with full routing tables covers only 126 ASes. All Tier-1 ASes are included in this set except AS209 (Qwest), but fortunately one of AS209's customer ASes hosts a monitor.

2) IXP Data: There are a number of websites, including Packet Clearing House (PCH) [8], Peeringdb [9], and Euro-IX [4], that maintain a list of IXPs worldwide together with a list of ISP participants in some IXPs. The list of IXP facilities is believed to be close to complete [10], but the list of ISP participants at the different IXPs is likely incomplete or outdated, since its input is done by the ISPs on a voluntary basis. However, most IXPs publish the subnet prefixes they use in their layer-2 clouds, and the best current practice [6] recommends that each IXP participant keeps reverse DNS entries for their assigned IP addresses inside the IXP subnet. Based on the above information, we adopted the method used in [43] to infer IXP participants. The basic idea is to do reverse DNS lookups on the IXP subnet IP addresses, and then infer the participating ISPs from the returned DNS names. From the aforementioned three data sources, we were able to derive a total of 6,084 unique presences corresponding to 2786 ASes in 204 IXPs worldwide. Table I shows the breakdown of the observed presences per data source. Note that a *presence* means that there exists an AS-IXP pair. For example, if two ASes peer at two IXPs, it will be counted as two presences. Although we do not expect our list to be complete, we noticed that the total number of presences we obtained is very close to the sum of the number of participants in each IXP disclosed on the PCH website.

3) IRR Data: The Internet Routing Registry (IRR) [7] is a database to register inter-AS connectivity and routing polices. Since registration with IRR is done by ISP operators on a voluntary basis, the data is known to be incomplete and many records are outdated. We filtered IRR records by ignoring all entries that had a "Last Modified" date that was more than one year old.

4) Proprietary Router Configurations and Syslogs: This is a major source for deriving the ground truth for our Tier-1 and Tier-2 ISPs, where the latter is a transit provider and a direct customer of the former. The data include historical configuration files of more than one thousand routers in these two networks, historical syslog files from all routers in the Tier-1 network, and "show ip bgp sum" outputs from all routers in the Tier-2 network. We also have access to iBGP feeds from several routers in these two networks.

5) Other Proprietary Data: To obtain the ground truth for other types of networks, we had conversations with the operators of a small number of content providers. Since large content providers are unwilling to disclose their connectivity information in general, in this paper we present a fictitious content provider whose numbers of AS neighbors, peer links, and IXP presences are consistent with the data we collected privately. We also obtained the ground truth of the AS-level connectivity for four stub networks from their operators.

D. Establishing the Ground Truth

We describe here the method we use to obtain the ground truth of AS level connectivity of the Tier-1 network; we use a similar process for the other networks. To obtain the AS-level connectivity ground truth, we need to know at each instant in time the BGP sessions that are in the established state for all the BGP routers in the network. A straightforward way to do this is to launch the command "show ip bgp summary" in all the routers simultaneously. Fig. 4 shows an example output produced by this command. The state of each BGP session can be inferred by looking at the column "State/PfxRcd" – when this column shows a numeric value, it refers to the number of prefixes received from the neighbor router, and it is implied that the BGP session is in *established* state. In this example, all connections are in the *established* state except for the session with neighbor 64.125.0.137, which is in the *idle* state.

Due to the large size of the Tier-1 network under study, it is infeasible to run the "show ip bgp sum" command over all the routers of the network and over a long study period. It is also impossible to obtain any historic "show ip bgp sum" data for the past. Therefore, we resort to an alternative way to infer the connectivity ground truth – analyzing routers' configuration files. Routers' configuration files are a valuable source of information about AS level connectivity. Before setting up a BGP session with a remote AS, each router needs to have a minimum configuration state. As an example, in Fig. 3, for router R_0 in AS10 to open a BGP session with R_2 in AS20, it needs to have a "neighbor 129.213.1.2 remote-as 20" entry in its configuration file, as well as IP connectivity between R_0 and R_2 through a configured route to reach R_2. Similarly, R_2 needs to have a configured route to reach R_0. The IP connectivity between the two routers of a BGP session can be established in one of the following two ways:

- **Single-hop**: two routers are physically connected directly, as the case of R_0 and R_2 in Fig. 3. More specifically R_0 can (1) define a *subnet* for the local interface at R_0 that includes the remote address 129.213.1.2 of R_2, e.g., " ip address 129.213.1.1 255.255.255.252" (where 255.255.255.252 is the subnet mask) or (2) set a *static route* in R_0 to the remote address 129.213.1.2 of R_2, e.g., "ip route 129.213.1.0 255.255.255.252 Serial4/1/1/24:0" (in this case, Serial4/1/1/24:0 refers to the name of the local interface at R_0).
- **Multihop**: two routers (such as R_1 and R_3 in Fig. 3) are not directly connected, but connected via other routers. To configure such a multihop BGP session, R_1 configures, e.g., "neighbor 175.220.1.2 ebgp-multihop 3" (here 3 refers to the number of IP hops between R_1 and R_3); R_1 reaches R_3 by doing longest prefix matching of 175.220.1.2 in its routing table.

```
Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
4.68.1.166      4  3356  387968    6706  1652742    0    0 4d15h      231606
64.71.255.61    4   812  600036    6706  1652742    0    0 4d15h      230964
64.125.0.137    4  6461       0       0        0    0    0 never    Idle
65.106.7.139    4  2828  466128    6706  1652742    0    0 4d15h      232036
```

Fig. 4. Output of "show ip bgp summary" command.

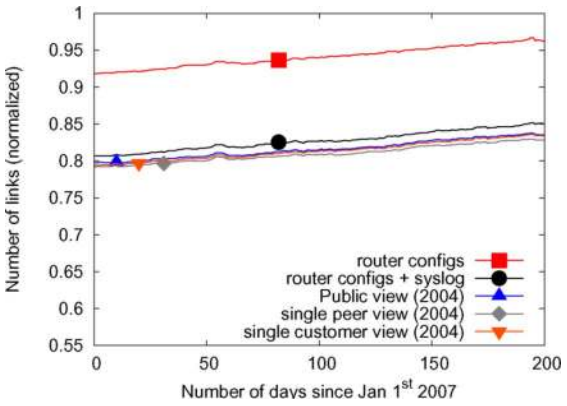


Fig. 5. Connectivity of the Tier-1 network (since 2004).

Ideally, we would like to verify the existence of a BGP session by checking the configuration files on both sides of a session. Unfortunately it is impossible to get the router configurations of the neighbor ASes. We thus limit ourselves to check only the configuration files of routers belonging to the Tier-1 network. We noticed that a number of entries in the router configuration files did not satisfy the minimal BGP configuration described above, probably because the sessions were already inactive, and these sessions should be discarded. After searching systematically through the historic archive of router configuration files, we ended up with a list of neighbor ASes that have at least one valid BGP configuration. The "router configs" curve in Fig. 5 shows the number of neighbor ASes in this list over time[2].

However, even after this filtering, we still noticed a considerable number of neighbor ASes that appeared to be "correctly configured," but did not have any established BGP session. This could be due to routers on the other side of the sessions not being configured correctly. Given that we do not have the configuration files for those neighbor routers, we utilize router syslog data to filter out the possible stale entries in the Tier-1's router configurations. Syslog records include information about BGP session failures and recoveries, indicating at which time each session comes up or goes down. More Specifically, a *BGP-5-ADJCHANGE* syslog message has the following format: "*timestamp local-router* BGP-5-ADJCHANGE: neighbor *remote-ip-address* Down", and it indicates the failure of the session between the *local-router* and the neighbor router whose IP address is *remote-ip-address*. We use the following two simple rules to further filter the previous list of neighbors:

1) If the last message of a session occurs at day t and the content was "session down", and there is no other message from the session in the period [t, $t+1$ month], then we assume the session was removed at day t (i.e., we wait at least one month before discarding the session).

[2]Note that the number is normalized for nondisclosure reasons.

2) If a session is seen in a router configuration at day t, but does not appear in syslog for the period [t, $t+1$ year], then we assume the session was removed at day t (i.e., we wait at least 1 year before discarding the session).

Note that the above thresholds were empirically selected to minimize the number of false positives and false negatives in the inferred ground truth. A smaller value would increase the number of false negatives (i.e., sessions that are prematurely removed by our scheme while still in the ground truth), whereas a higher value would increase the false positives (i.e., sessions that are no longer in the ground truth, but have not been removed yet by our scheme). We calibrated the thresholds using AS adjacencies that were present in both the syslog messages and in the public view, e.g., we quantified the false negatives by looking at adjacencies that we excluded using the syslog thresholds, but were actually still visible in the public view. Even though these threshold values worked well in this case, depending on the stability of links and routers' configuration state, other networks may require different values. Note also that these two rules are for individual BGP sessions only. An AS-level link between the Tier-1 ISP and a neighbor AS will be removed only when *all* of the sessions between them are removed by the above two rules. The sessions between the Tier-1 ISP and its peers tend to be stable with infrequent session failures [41], thus it is possible that a session never fails within a year. But our second rule above is unlikely to remove the AS-level link between the Tier-1 ISP and its peer because there are usually multiple BGP sessions between them and the probability that none of the sessions have any failures for an entire year is very small. Similarly, this argument is true for large customer networks which have multiple BGP sessions with the Tier-1 ISP. On the other hand, small customers tend to have a small number of sessions with the Tier-1 ISP (perhaps one or two), and the sessions tend to be less stable thus have more failures and recoveries. Thus if the AS link exists, the above two rules should not filter it out since some syslog session up or down messages will be seen. For similar reasons, the results are not significantly affected by the fact that some syslog messages might be lost in transmission due to unreliable transport protocol (UDP). Using the two simple rules above, we removed a considerable number of entries from the config files, and obtained the curve "router configs+syslog" in Fig. 5; note that our measurement started in 2006-01-01, but we used an initial 1-year window to apply the second syslog rule. In the next section we compare in detail the inferred ground truth with the observable connectivity in the public view for different networks, including the Tier-1.

III. Case Studies

In this section, we compare the ground truth of networks for which we have operational data with the connectivity derived from the public view to find out what links are missing from the latter and why they are missing.

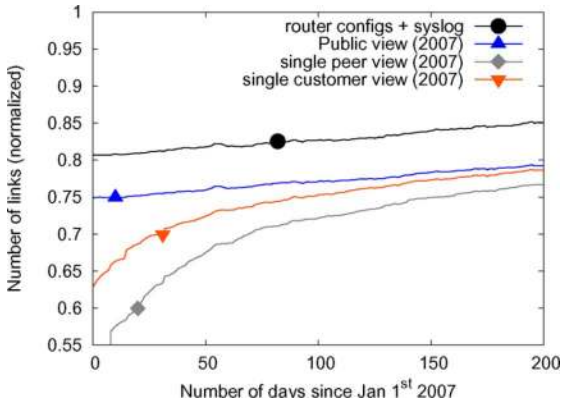


Fig. 6. Connectivity of the Tier-1 network (since 2007).

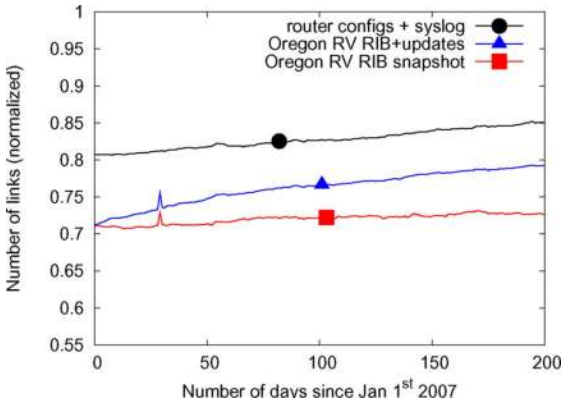


Fig. 7. Capturing the connectivity of the Tier-1 network through table snapshots and updates.

A. *Tier-1 Network*

Once we achieved a good approximation of the ground truth as described in the previous section, we compared it to the public view derived connectivity. For each day t, we compared the list of ASes in the inferred ground truth $T_{\text{tier1}}(t)$ obtained from router configs+syslog, with the list of ASes seen in public view as connected to the Tier-1 network up to day t. The "Public view (2004)" curve in Fig. 5 is obtained by accumulating public view BGP-derived connectivity since 2004. We first note that *all* the Tier-1 ISP's links to its peers and sibling ASes are captured by the public view. In particular, we note that the public view captured all the peer-peer links of the Tier-1 ISP. The peer links of an AS are visible as long as a monitor resides in the AS itself, or in any of the AS's customers, or the customer's customers. In fact, the public view captured all the peer-peer links for all Tier-1 ASes, due to the small number of Tier-1 networks and the fairly large set of monitors used by public view.

Comparing the "Public view (2004)" curve with the "router configs+syslog" curve in Fig. 5, we also note that there is an almost constant small gap, which is of the order of some tens of links (3% of the total links in "router configs+syslog"). We manually investigated these links, and found that there are three main causes for why they do not show up in the public view: 1) the links that connect to the Tier-1's customer ASes which only advertise prefixes longer than /24; these long prefixes are then aggregated by the Tier-1 AS before announcing to other neighbors. This category accounts for about half of the missing links; 2) there is one special purpose AS number (owned by the Tier-1 ISP) which is only used by the Tier-1 ISP; 3) false positives, i.e., ASes that were wrongly inferred as belonging to $T_{\text{tier1}}(t)$, including stale entries, as well as newly allocated ASes whose sessions were not up yet. The false positive contributes to about half of the "missing links" (which should not be called "missing").

Fig. 6 shows similar curves using the same vertical scale as in Fig. 5, but this time the public view BGP data collection is started in the beginning of 2007. When comparing "Public view (2007)" and "router configs+syslog" we note the gap is bigger, indicating that some entries in "router configs+syslog" did not show up in public view after 2007, but they did show up before, which likely means they are stale entries (false positives).

The "Single customer view" and "Single peer view" curves in both Figs. 5 and 6 represent the Tier-1 connectivity as seen from a single router in a customer of the Tier-1 ISP and a single router in a peer of the ISP, both from the public view. The single peer view captures slightly less links than the single customer view, corresponding to about $\sim 1.5\%$ of the total number of links of the Tier-1 network. Further analysis revealed that this small delta corresponds to the peer links of the Tier-1, which are included in routes advertised to the customer but *not* advertised to the peer. This is expected and consistent with the no-valley routing policy. We also note that the "Single peer view" and "Single customer view" curves in Fig. 6 show an exponential increase in the first few days of the x axis, which is caused by the revelation of hidden links, as explained in Section II-B. However, the nine months of the measurement should be enough to reveal the majority of the hidden links [33]. In addition, note that in both figures, the "Single customer view" curve is very close to the public view curve, which means that the connectivity of the Tier-1 as seen by the customer is representative of what is visible from the public view.

Fig. 7 shows the difference between using routing table snapshots (RIB) versus using an initial RIB *plus* BGP updates from all the routers at Oregon RouteViews (a subset of 46 routers of the entire public view). Note that on each day, the number of links in the curves "Oregon RV (RouteViews) RIB snapshot" and "Oregon RV RIB+updates" represent the overlap with the set of links in the inferred ground truth represented by the curve "router configs+syslog," i.e., those links not in "router configs + syslog" are removed from the two "Oregon RV" curves. Even though both curves start in the same point, after more than nine months of measurement, "Oregon RV RIB+upates" reveals about 10% more links than those revealed by "Oregon RV RIB snapshot," these are the links that were revealed in BGP updates of alternative routes encountered during path exploration as described in [32]. We also note that the difference between the two curves are all customer-provider links, and all the Tier-1 ISP's links to the peers are captured by the "Oregon RV RIB snapshot," because of the large number of routes that go through these peer-peer links.

1) Summary:

- A single snapshot of the Oregon RV RIB can miss a noticeable percentage (e.g., 10%) of the Tier-1's AS-level links, all of them customer-provider links, when compared to using RIBs plus updates accumulated in several months.
- The Tier-1 AS's links are covered fairly completely by the public view over time. All the peer-peer and sibling

links are covered; the small percentage (e.g., 1.5%) of links missing from public view are the links to customer ASes who only announce prefixes longer than /24 and hence their routes are aggregated.

- The Tier-1 AS's links are covered fairly completely by a single customer by using the historic BGP tables and updates, which can be considered representative of the public view.
- The Tier-1 AS's links are covered fairly completely by a single peer (when the historic BGP table and updates are used), and the about 1.5% missing links are all peer-peer links.

B. Tier-2 Network

The Tier-2 network we studied differs from the previous Tier-1 case in a few important ways. First of all, not being a Tier-1 network, the Tier-2 has providers. Second, it is considerably smaller in size as measured by the number of routers, however it has considerably more peer links than the Tier-1 network. Third, the Tier-1 network peers exclusively through private peering, this Tier-2 network had close to $\frac{2}{3}$ of its peers through IXPs. We do an analysis similar to the Tier-1 case, except that we did not have access to syslog data.

The "router configs" curve in Fig. 8 shows the number of neighbor ASes obtained from router configurations over time. Let us assume for now this is a good approximation of the ground truth of the Tier-2 network connectivity. We include in Fig. 2 two single router view curves, one is obtained from a router in a customer of the Tier-2 network, and the other is derived from a router in a provider of the Tier-2 network, both are in the public view. Note that this time we started the measurement in March 2007 when the BGP data for the customer router became available in the public view. This customer router became unavailable after August 13, 2007, hence, the single customer view curve is chopped off after that date. Fig. 8 shows that the provider view misses a significant number of links that are captured by the customer view. This difference amounts to more than 12% of the Tier-2's links captured by the customer, which are all the peer links of the Tier-2 network. For comparison, we also included the public view curve, starting March 10, 2007. Note that the public view captured a very small number of neighbors that are not in the customer view. We found that most of the links in this small gap were revealed in the routes that were originated by the Tier-2's customers and had several levels of AS prepending. The customer we used for the customer view curve did not pick these routes because of the path inflation due to the AS prepending, however following the prefer-customer policy, routers in the Tier-2 network picked these prepended routes, and one of these routers is in the public view data set.

From Fig. 8 we also note that the connectivity captured by the public view is $\sim 85\%$ of that inferred from router configs, which could be due to incorrect or stale entries in the router configuration files. To verify whether this is the case, we launched a "show ip bgp summary" command on all the routers of the network on 2007-09-03, and we take into account only those BGP sessions that were in the established state. The number of neighbors with at least one such session is shown in Fig. 8 by the "show ip bgp sum" point, which has only 80% of the connectivity inferred

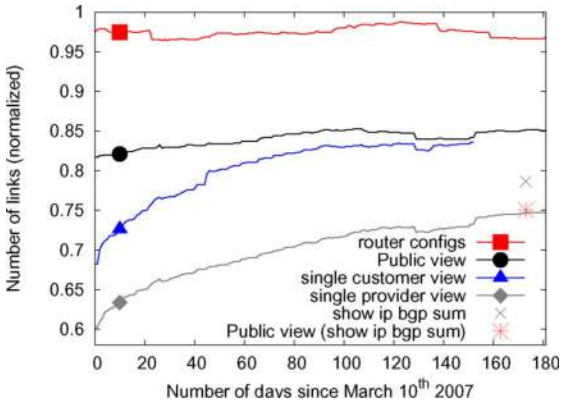


Fig. 8. Tier-2 network connectivity.

from the router configurations. This means that about 20% of the connectivity extracted from router configs were false positives. On the other hand, we observe that by accumulating BGP updates over time, we also increase the number of false positives, i.e., adjacencies that were active in the past and became inactive. By comparing the curves "Public View" and "Public view (show ip bgp sum)," we note that about $1-\frac{0.75}{0.85} \simeq 0.12$ (or 12%) of the links accumulated in public view over the 6-month period correspond to false positives. There are however ways to filter these false positives: (1) by removing the short-lived links, since most likely they correspond to misconfigurations, or (2) by timing out links after a certain period of time. The point "Public view (show ip bgp sum)" in the figure represents the intersection between the set of neighbors extracted from "show ip bgp sum" and the set of neighbors seen so far in the public view. Note that public view missed $\sim 7\%$ of the links given by "show ip bgp sum," which amounts to a few tens of links. One of these links was the RouteViews passive monitoring feed, some other were internal AS numbers, and the remaining ones were to the ASes announcing longer than /24 routes (that were aggregated). Note also that the fairly complete coverage of the Tier-2 network's connectivity is due to the existence of a monitor residing in a customer of the Tier-2. As we explained in the Tier-1's study, the public view can capture all the links, including all peer links of an AS, if a monitor resides in either the AS itself, or in the AS's customer or customer's customers.

Fig. 9 shows the difference between using single RIB snapshot versus initial RIB+updates from RouteViews Oregon collector, using the same vertical scale as in Fig. 8. In this case, using updates reveals $\sim 12\%$ more links than those revealed by router RIB snapshots in the long run. Note that there is a lack of configuration files at beginning of 2007, hence, the missing initial part on the curve "router configs." The jump in the figure is due to the addition of the monitor in the Tier-2 customer AS, which revealed the peer links of the Tier-2 network.

1) Summary:

- A single snapshot of the Oregon RV RIB can miss a noticeable percentage (e.g., 12%) of the Tier-2's AS-level links, all of them customer-provider links, when compared to using RIBs+updates accumulated in several months.
- The Tier-2 AS's links are covered fairly completely by a single customer over time (RIBs +updates), which can be considered representative of the entire public view.

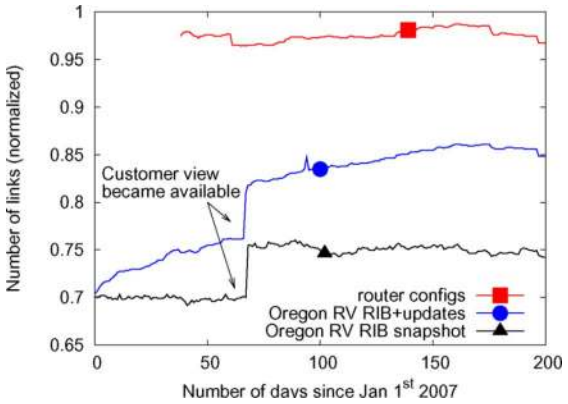


Fig. 9. Capturing Tier-2 network connectivity through table snapshots and updates.

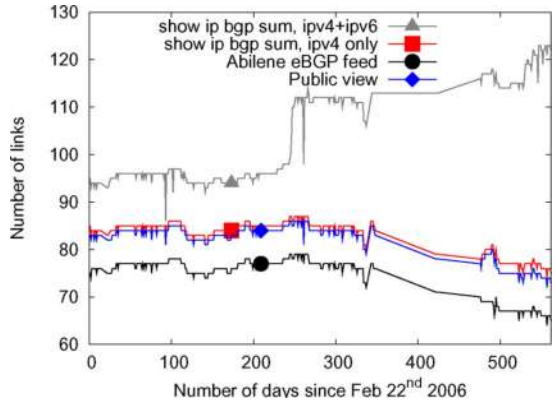


Fig. 10. Abilene connectivity.

- A single provider view can miss a noticeable percentage (e.g., 12%) of the Tier-2's links, and all the missing links are peer-peer links.
- A Tier-2 AS's links are covered fairly completely by the public view over time if there is a monitor in it, or its customer or its customer's customers, in which case all the peer-peer links are revealed. The small percentage (e.g., 7%) of links missing from the public view are those connecting to customers who only announce prefixes longer than /24 or those ASes dedicated for internal use.

C. Abilene and Geant

1) Abilene: Abilene (AS11537) is the network interconnecting universities and research institutions in the US. The Abilene Observatory [14] keeps archives of the output of "show ip bgp summary" for all the routers in the network. Using this data set, we built a list of Abilene AS neighbors over time, which is shown in the "show ip bgp sum, ipv4+ipv6" curve in Fig. 10. Even though Abilene does not provide commercial transit, it enables special arrangements where its customers may inject their prefixes to commercial providers through Abilene, and receive routes from commercial providers through Abilene. The academic-to-commercial service is called Commercial Peering Service (or CPS) versus the default academic-to-academic Research & Education (R&E) service. These two services are implemented by two different VPNs over the Abilene backbone. BGP sessions for both VPNs are included in the output of "show ip bgp summary." We compare Abilene connectivity ground truth with that derived from a single router eBGP feed (residing in Abilene) containing only the R&E sessions. In addition, we do a similar comparison with our public view, which should contain both CPS and R&E sessions (public view contains eBGP+iBGP Abilene feeds, as well as BGP data from commercial providers of Abilene). However, since there are a considerable number of neighbors in Abilene that are using IPv6 only, and since the BGP data in our data set are mostly IPv4-only, we decided to place the IPv4-only neighbors in a separate set. The curve "show ip bgp sum, ipv4 only" in Fig. 10 shows only the AS neighbors that have at least one IPv4 session connected to Abilene[3]. Contrary to the "show ip bgp sum, ipv4+ipv6" curve which includes all sessions, the IPv4-only curve shows a decreasing trend. We believe this is because some of the IPv4 neighbors have been migrating to IPv6 over time. When comparing the "show ip bgp sum, ipv4 only" curve with the one derived from the eBGP feed, we find there is a constant gap of about 10 neighbors. A closer look into these cases revealed that these AS numbers belonged to commercial ASes with sessions associated with the CPS service. The small gap between the public view and the IPv4-only curve corresponds to the passive monitoring session with RouteViews (AS6447).

2) Geant: Geant (AS20965) is an European research network connecting 26 R&E networks representing 30 countries across Europe. In contrast to Abilene where the focus is on establishing academic-to-academic connectivity, Geant enables its members to connect to the commercial Internet using its backbone. We inferred Geant connectivity ground truth by running the command "show ip bgp sum" in all its routers through its looking glass site [5]. We found a total of 50 AS neighbors with at least one session in the established state. By comparing Geant ground truth with the connectivity revealed in public view, we found a match on all neighbor ASes except two. One of the exceptions was a neighbor which was running only IPv6 multicast sessions, and therefore hidden from public view which consists mostly of IPv4-only feeds. The other exception seems due to a passive monitoring session to a remote site, which explains why its AS number was missing from BGP feeds.

3) Summary: In Abilene and Geant, the public view matches the connectivity ground truth (no invisible or hidden links), capturing all the customer-provider and peer links. Abilene represents a special case, where depending on the viewpoint there can be invisible links. For instance, some Abilene connectivity may be invisible to its customers due to the academic-to-commercial special arrangements.

D. Content Provider

Content networks are fundamentally different from transit providers such as the Tier-1 and Tier-2 cases we studied earlier. Content networks are edge ASes and do not transit traffic between networks, thus they only have peers and providers. They generally try to reduce the amount of (more expensive) traffic sent to providers by directly peering with as many other networks as possible; direct peerings can also help improve performance. Consequently, content networks in general have a heavy presence at IXPs, where they can peer with multiple different

[3]Note that there was a period of time between days 350 and 475 for which there was no "show ip bgp sum" data from Abilene.

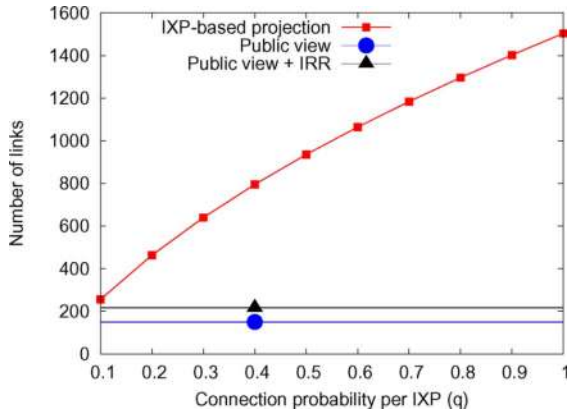


Fig. 11. Projection of the number of peer ASes of a representative content provider.

networks. While two transit providers usually peer at every location where they have a common presence in order to disperse traffic to closer exit-points, peering of content networks is more "data-driven" (versus "route-driven"), and may happen in only a fraction of the IXPs where two networks have common locations. Based on this last observation, we estimate the connectivity of a representative content provider C, and compare it to the connectivity observed from the public view. We assume that in each IXP where C has presence, it connects to a fixed fraction q of the networks that are also present at that IXP, i.e., if C has n common locations with another network X, 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 is summed over all the 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 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. Furthermore, we know that the number of providers of C is negligible compared to the number of its peers, and that more than 95% of its peerings are at IXPs. Therefore it is reasonable to represent the AS-level connectivity of C by its peerings at IXPs.

Fig. 11 shows the projection of the number of neighbor ASes of C as a function of the connection probability q at each IXP. For comparison purposes, we also include the number of neighbor ASes of C as inferred from the public view over a window of 6 months. From discussions with C's operators, we know that at each IXP, C peers with about 80%-95% of the participants at the IXP (parameter q), and that the total number of BGP sessions of C is more than 2000, even though we do not know the total number of unique peer ASes[4]. In view of these numbers, the projection in Fig. 11 seems reasonable, even after taking into account that our IXP membership data is incomplete. The most striking observation is the amount of connectivity missed from the public view, which is on the order of thousands of links and represents about 90% of C's connectivity. This result is not entirely surprising, however, because based on no-valley policy, the content provider C does not announce its peer-peer links to anyone, and a peer-peer link is visible only if the public view has a monitor in C, or in the peer or a customer of the peer. Yet the number of available monitors is much smaller than the projected total number of C's peer. We believe this result holds true for other large content providers, search engines, and content distribution networks.

Trying to close the gap between reality and the public view, we looked for additional connectivity in the IRR, as described in Section II-C. We discovered 62 additional neighbor ASes for C that were not present in the initial set of 155 ASes seen in the public view. Even though this addition increased the number of covered neighbor ASes of C to 217, it is still only about 15% of the AS-level connectivity of C.

1) Summary: The public view misses about 90% of C's connectivity, and we believe all of them are invisible peer-peer links, and most of them are likely at IXPs. Using IRR information reduces the missing connectivity slightly, to 85%. The public BGP view's inability to catch these peer-peer links is due to the no-valley policy and the absence of monitors in the peers or their customers of the content network.

TABLE II
CONNECTIVITY OF STUB NETWORKS

Network	# of neighbor ASes in ground truth	#of neighbor ASes in public view
A	8	10
B	7	6
C	3	4
D	2	2

E. Simple Stubs

Stub networks are those ASes that do not have customers (or have a very small number of customers)[5]. Stubs represent the vast majority of ASes, and they are typically sorted according to their business rationale into: 1) content, 2) eyeball, and 3) simple. Content networks have heavy outbound traffic, whereas eyeballs are heavy inbound (e.g., cable/dsl providers). Simple stubs represent enterprise customers such as universities and small companies. We obtained the AS-level connectivity ground truth of 4 simple stubs by directly contacting their operators. Table II shows for each network the number of neighbor ASes in the ground truth as reported by the operators, as well as the number of neighbor ASes captured by the BGP-derived public view. Note that for public view we use 6 month worth of BGP RIB and updates to accumulate the topology to account for hidden links that take time to be revealed [33]. Network D is the only case where there is a perfect match between ground truth and public view. For network A, there are two neighbors included in public view that were disconnected during the 6-month window (false positives). For network B, the public view was missing a neighbor due to a special agreement in which the routes learned from the neighbor are not announced to B's provider. Finally, for network C there was an extra neighbor in public view that was never connected to C, but appeared in routes during one day in the 6-month window. We believe this case was originated either by a misconfiguration or a malicious false link attack.

1) Summary: The 6-mo accumulated public view captured all the customer-provider links of the stub networks studied. In total, the public view has one false negative (invisible link) and

[4]The number of unique neighbor ASes is less than the total number of BGP sessions, as there exist multiple BGP sessions with the same neighbor AS.

[5]The details about stub classification are describe in Section IV-B

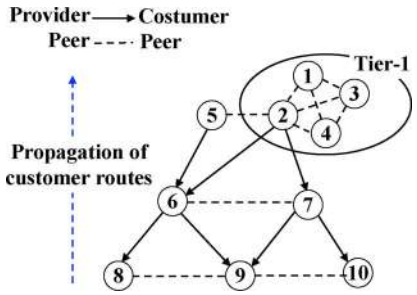


Fig. 12. Customer-provider links can be revealed over time, but downstream peer links are invisible to upstream monitors.

3 false positives, the latter can be eliminated by reducing the interval of the observation window of public view.

IV. Completeness of the Public View

In this section, we first summarize the classes of topological information that are captured and necessarily missed in the public view. Based on this observation, we then describe a novel method to infer the business relationships between ASes. We use the inferred relationships to do AS classification and determine how much of the topology is covered by the current set of monitors in the public view.

A. "Public View" Versus Ground Truth

We use Fig. 12 as an illustration to summarize the degree of completeness of the observed topology as seen by the public view. Our observations presented here are the natural results of the *no-valley-and-prefer-customer* policy, and some of them have been speculated briefly in previous work. In this paper, we quantify and verify the degree of completeness by comparing the ground truth with the observed topology. Though the few classes of networks we have examined are not necessarily exhaustive, we believe the observations drawn from these case studies provide insights that are valid for the Internet as a whole.

First, if a monitor resides in an AS A, the public view should be able to capture all of A's direct links, including both customer-provider and peer links. However, not all the links of the AS may show up in a snapshot observation. It takes time, which may be as long as a few years, to have all hidden customer-provider links exposed by routing dynamics. Second, a monitor in a provider network should be able to capture all the provider-customer links between itself and all of its downstream customers, and a monitor in a customer network should be able to capture all the customer-provider links between itself and its upstream providers. For example, in Fig. 12, a monitor in AS2 can capture not only its direct provider-customer links (2-6 and 2-7), but also the provider-customer links between its downstream customers (6-8, 6-9, 7-9, and 7-10). AS5, as a peer of AS2, is also able to capture all the provider-customer links downstream of AS2 since AS2 will announce its customer routes to its peers. Again, it can take quite a long time to reveal all the hidden links. Third, a monitor cannot observe a peer link of its customer, or peer links of its neighbors at the same tier[6]. For example, a monitor at AS5 will not be able to capture the peer link 6-7 or 1-2, because a peer route is not announced to providers or other peers according to the *no-valley* policy. Fourth, to capture a peer link requires a monitor in one of the peer ASes or in a downstream customer of the two ASes incident to the link. For example, a monitor at AS9 can observe the peer links 6-7 and 5-2, but not the peer link 1-3 since AS9 is not a downstream customer of either AS1 or AS3.

The current public view has monitors in all the Tier-1 ASes except one, and that particular Tier-1 AS has a direct customer AS that hosts a monitor. Applying the above observations, we can summarize and generalize the completeness of the AS-level topology captured by the public view as follows.

- **Coverage of Tier-1 links**: The public view contains all the links of all the Tier-1 ASes.
- **Coverage of customer-provider links**: There is no invisible customer-provider link. Thus over time the public view should be able to reveal all the customer-provider links in the Internet topology, i.e., the number of hidden customer-provider links should gradually approach zero with the increase of the observation period length. This is supported by our empirical findings: in all our case studies we found all the customer-provider links from BGP data collected over a few years.
- **Coverage of peer links**: The public view misses a large number of peer links, especially peer links between lower tier ASes in the Internet routing hierarchy. The public view will not capture a peer link A–B unless there is a monitor installed in either A or B, or in a downstream customer of A or B. Presently, the public monitors are in about 400+ ASes out of a total over 27 000 existing ASes, this ratio gives a rough perspective on the percentage of peer links missing from the public view. Peer links between stub networks (i.e., links 8-9 and 9-10 in Fig. 12) are among the most difficult ones to capture. Unfortunately, with the recent growth of content networks, it is precisely these links that are rapidly increasing in numbers.

B. Network Classification

The observations from the last section led us to a novel and simple method for inferring the business relationships between ASes, that allow us also to classify ASes in different types.

1) Inferring AS Relationships: The last section concluded that, assuming routes follow a no-valley policy, monitors at the top of the routing hierarchy (i.e., those in Tier-1 ASes) are able to reveal all the downstream provider-customer connectivity over time. This is an important observation since, by definition, each non-Tier-1 AS is a customer of at least one Tier-1 AS, then essentially all the provider-customer links in the topology can be observed by the Tier-1 monitors over time. This is the basic idea of our AS relationship inference algorithm.

We start with the assumption that the set of Tier-1 ASes is already known[7]. By definition of Tier-1 ASes, all links between Tier-1s 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-$\cdots$-a_n. The link m-a_1 can be either a provider-customer link, or a peer link (this is because in certain cases a Tier-1 may 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, $\ldots$, a_{n-1}-a_n must be provider-customer links, because a peer or provider route should not be propagated

[6]We assume that the provider-customer links do not form a circle.

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

upstream from a_1 to m. Therefore, the segment $a_2, \ldots, a_n$ must correspond to a customer route received by a_1. To infer the relationship of m-a_1, we note that according to no-valley policy, if m-a_1 is a provider-customer link, this link should appear in the routes propagated from m to other Tier-1 ASes, whose monitors will reveal this link. On the other hand, if m-a_1 is a peer link, it should never appear in the routes received by the monitors in other Tier-1 ASes. Given we have monitors in all Tier-1 ASes or their customer ASes, we can accurately infer the relationship m-a_1 by examining whether it is revealed by other Tier-1 ASes. Using this method, we can first find and label all the provider-customer links, and then label all the other links revealed by the monitors as peer links.

Our algorithm is illustrated in Fig. 12, where 1, 2, 3, and 4 are known to be Tier-1s. Suppose AS 2 monitor reveals an ASPATH 2-5-6-8 and another ASPATH 2-7-9; while monitors at AS 4 reveals an ASAPTH 4-2-7-9, but none of 1, 3, 4 reveals an AS-PATH with segment of 2-5-6-8. According to our new method, 5-6, 6-8, and 7-9 are definitely provider-customer links. 2-7 is a provider-customer link since it is revealed by Tier-1s other than 2, while 2-5 is a peer link since it is not revealed by any other Tier-1s. Furthermore, suppose AS 6 is a monitor and it reveals link 6-7, and 6-7 is never revealed by Tier-1 ASes 1,2,3, or 4. Then we can conclude that this 6-7 is a peer link.

From BGP data collected from all the Tier-1 monitors over a 7-mo period, we were able to infer a total of 70,698 provider-customer links. We also noticed that a small number of these links only existed 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 route hijacks, as described in [27]. After filtering all the routes with a lifetime less than 2 d over the 7-mo measurement period, we excluded 5,239 links, ending up with a total of 65 459 provider-customer links. Note that even though our relationship inference has the advantage of being simple, its accuracy can still be improved. For instance, we could use the algorithm in [23] to select a maximal set of AS paths that do not create cycles in relationships and are valley-free, and only consider such relationships as valid. Note that out algorithm differs from the classic Gao's algorithm [22] in several ways. First, our algorithm is able to infer all the customer provider relations based only in a very limited number of sources (the Tier-1 routers). Second, contrary to [22], we do not rely on node degree to infer peer relationships. In fact, the node degree is a variable of the monitor set, and that is the main reason why [22] produces so distinct results with varying monitor sets. Our inference of peer relationships is purely based on the no-valley premise that peer routes are not propagated upstream, therefore we believe our inference results are more accurate.

2) AS Classification: AS classification schemes are typically based on each AS's node degree (the number of neighbors) or the number of prefixes originated. However, the degree can be misleading since it is a mix of providers, peers and customers in one count, and the number of prefixes originated is not very reliable either since the length of the prefixes is different and the routes carried downstream are not accounted. With the inferred provider-customer relations in hand, we decided to use the number of downstream customer ASes (or "customer cone") as also defined in [20]. Fig. 13 shows the distribution of the number of downstream customers per AS. We note that over

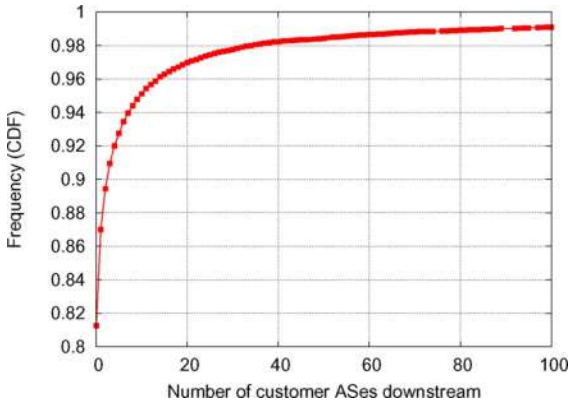


Fig. 13. Distribution of number of downstream customers per AS.

TABLE III
COVERAGE OF BGP MONITORS

Parameter	Full tables	Full+partial tables
No. monitored ASes	121	411
Covered ASes	1,101 / 28,486 $\simeq$ 4%	1,552 / 28,486 $\simeq$ 5 %

TABLE IV
COVERAGE OF BGP MONITORS FOR DIFFERENT NETWORK TYPES

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

80% of the ASes have no customers, and a noticeable fraction of ASes have a very small number of customers. We label as *stub* those ASes with 4 or less customers, which encompass about 92% of the ASes. This should correspond to end networks which either don't provide transit or have very limited transit to few local customers, e.g., universities providing transit to small local research facilities. Based on the knee of the distribution in Fig. 13, we label as *small ISPs* those ASes with between 5 and 50 downstream customers. They correspond to about 6% of the total ASes. The remaining non-Tier-1 ASes in the long tail are labeled as *large ISPs*. Table IV shows the number of ASes in each class. We analyzed the sensitivity of the classification thresholds by changing their values by some delta, and did not notice significant difference in the end result.

C. Coverage of the Public View

With our new method for AS relationship inference and AS classification, we now attempt a rough quantification of the completeness of the AS topology as observed by the public view. According to our observations in Section IV-A, a monitor can uncover all the upstream connectivity over time. For example, in Fig. 12, a monitor at AS 7 will receive routes from upstream providers that will carry the peer links existing upstream, in this case the links 2-1, 2-3, 2-4 and 2-5 (in addition to the upstream provider-customer links). Therefore, by starting at AS 7 and following all provider-customer links upstream, we pass through all the ASes that are *covered* by a monitor in AS 7, in the sense that this monitor is able to reveal all their

connectivity. In Fig. 12, AS 7 only covers AS 2, but AS 9 covers 4 upstream ASes: 6, 7, 2, and 5.

We applied this reasoning to the monitored ASes in the public view, and the results are shown in Table III. For comparison purposes, we included the results from using the set of monitors with full routing tables and that from using all the monitors with either full or partial routing tables; the difference between the two sets is small. Among the 400+ monitors, only a minority have full tables, and due to the overlap in covered ASes between Routeviews and RIPE-RIS, the set of monitors with full tables correspond to only 126 ASes. This set of monitors in the public view is only able to cover 4% of the total number of ASes in Internet. This result indicates that the AS topologies derived from the public view, which have been widely used by the research community, may miss most of the peer connectivity within the remaining 96% of the ASes (or 57% of the transits).

Finally, we look at the covered ASes in terms of their classes, which is shown in Table IV. The column "Covered ASes-aggregated" refers to the fraction of covered ASes in each AS class, whereas the column "Covered ASes-by covering type" refers to the total 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 a total of 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%). These results are due to the fact that most of the monitors reside in the core of the network. In order to cover a stub, we would need to place a monitor in that stub, which is infeasible due to the very large number of stubs in Internet.

V. Discussion

The defects in the inferred AS topologies, as revealed by our case studies, may have different impacts on the different research projects and studies that use an inferred AS topology. In the following, we use a few specific examples to illustrate some of the problems that can arise.

A. Stub AS Growth Rates and Network Diameter

Given that the public view captures almost all the AS nodes and customer-provider links, it provides an adequate data source for studies on AS-topology metrics including network diameter; growth rates and trends for the number of stub ASes; and quantifying customer multihoming (where multihoming here does not account peer links).

B. Other Graph-Theoretic Metrics

Given that the public view is largely inadequate in covering peer links, and given that these peer links typically allow for shortcuts in the data plane, relying on the public view can clearly cause major distortions when studying generic graph properties such as node degrees, path lengths, node clustering, etc.

C. Impact of Prefix Hijacking

Prefix hijacking is a serious security threat facing Internet and happens when an AS announces prefixes that belong to other ASes. Recent work on this topic [25], [42], [15], [45] evaluates the proposed solutions by using the inferred AS topologies from the public view. Depending on the exact hijack scenario, an incomplete topology can lead to either an underestimate or overestimate of the hijack impact. Fig. 14 shows an example of a hijack simulation scenario, where AS2 announces prefix p that belongs to AS1. Because of the invisible peer link 1-2, the number of impacted ASes is underestimated, i.e., ASes 3,5 and 6 are believed to pick the route originated by AS1, whereas in reality they would pick the more preferred peer route coming from the hijacker AS2. At the same time, an incomplete topology could also lead simulations to overestimate the impact of a hijack. For example, the content network C considered in Section III has a large number of direct peers who are unlikely to be impacted by a hijack from a remote AS, so missing 90% of C's peer links in the topology would significantly overestimate the impact of such a hijack. On the other hand, if C is a hijacker, then the incomplete topology would result in a vast underestimation of the impact.

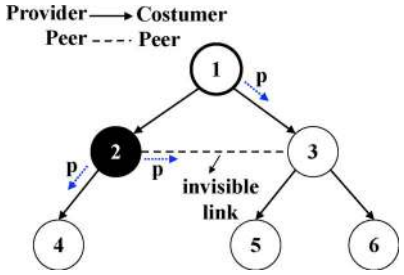


Fig. 14. Example of a prefix hijack scenario where AS2 announces prefix p belonging to AS1. Because of the invisible peer link AS2–AS3, the number of ASes affected by the attack is underestimated.

D. Relationship Inference/Path Inference

Several studies have addressed the problem of inferring the relationship between ASes based on observed routing paths [22], [39], [28]. There can be cases where customer-provider links are wrongly inferred as peer links based on the observed set of paths, creating a no-valley violation. Knowledge of the invisible peer links in paths could avoid some of these errors. The path inference heuristics [28], [30], [31] are also impacted by the incompleteness problem, mainly because they a priori exclude all paths that traverse invisible peer links.

E. Routing Resiliency to Failures

Studies that address robustness properties of the Internet under different failure scenarios (e.g., see [21], [42]) also heavily depend on having a complete and accurate AS-level topology, on top of which failures are simulated. One can easily envision scenarios where two parts of the network are thought to become disconnected after a failure, while in reality there are invisible peer links connecting them. Given that currently inferred AS maps tend to miss a substantial number of peer links, robustness-related claims based these inferred maps need to be viewed with a grain of salt.

F. Evaluation of New Interdomain Protocols

The evaluation of new interdomain routing protocols also heavily relies on the accuracy of the AS-level topology over which a new protocol is supposed to run. For instance, [40] proposes a new protocol where a path-vector protocol is used

among Tier-1 ASes, and all the ASes under each Tier-1 run link-state routing. The design is based on an assumption that customer trees of Tier-1 ASes are largely disjoint, and violations of this assumption are handled as rare exceptions. However, in view of our findings, there are a substantial number of invisible peer links interconnecting ASes at lower tier and around the edge of Internet, therefore connectivity between different customer trees becomes the rule rather than the exception. We would imagine the performance of the proposed protocol under complete and incomplete topologies to be different, possibly quite significantly.

VI. Related Work

Three main types of data sets have been available for AS-level topology inference: (1) BGP tables and updates, (2) traceroute measurements, and (3) Internet Routing Registry (IRR) information. BGP tables and updates have been collected by the University of Oregon RouteViews project [12] and by RIPE-RIS in Europe [11]. Traceroute-based datasets have been gathered by CAIDA as part of the Skitter project [13], by an EU-project called Dimes [37], and more recently by the iPlane project [26]. Other efforts have extended the above measurements by including data from the Internet Routing Registry [17], [38], [43]. However, studies that have critically relied on these topology measurements have rarely examined the data quality in detail, thus the (in)sensitivity of the results and claims to the known or suspected deficiencies in the measurements has largely gone unnoticed.

Chang *et al.* [17], [19], [16] were among the first to study the completeness of commonly used BGP-derived topology maps; later studies [44], [34], [43], using different data sources, yielded similar results confirming that 40% or more AS links may exist in the actual Internet but are missed by the BGP-derived AS maps. He *et al.* [43] report an additional 300% of peer links in IRR compared to those extracted from BGP data, however this percentage is likely inflated since they only took RIB snapshots from 35 of the $\sim$ 700 routers providing BGP feeds to RouteViews and RIPE-RIS. All these efforts have in common that they try to *incrementally* close the completeness gap, without first quantifying the degree of (in)completeness of currently inferred AS maps. Our paper relies on the ground truth of AS-level connectivity of different types of ASes to shed light on *what* and *how much* is missing from the commonly used AS maps and *why*. Dimitropoulos *et al.* [20] use AS adjacencies as reported by several ISPs to validate an AS relationship inference heuristic. They found that most links reported by ISPs that are not in the public view are peer links. In contrast to their work, most of our findings are inferred from iBGP tables, router configs, and syslog records collected over time from thousands of routers. Our approach yields an accurate picture of the ground truth as far as BGP adjacencies are concerned and allows us to verify precisely for each AS link x, why x was missing from public view. Last, in view of the recent work [36] that concludes that an estimated 700 route monitors would suffice to see 99.9% of all AS-links, our approach shows that such an overall estimate comes with an important qualifier: what is important is not the total number of monitors, but their locations within the AS hierarchy. In fact, our findings suggest a simple strategy for placing monitors to uncover the bulk of missing links, but unfortunately researchers have in general little input when it comes to the actual placement of new monitors.

VII. Conclusion

In this paper, we demonstrated the infeasibility to obtain a complete AS-level topology through the current data collection efforts, a direction that has been pursued in the past. We also attacked the problem from a new and different angle: obtaining the ground truth of a sample set of ASes' connectivity structures and comparing them with the AS connectivity inferred from publicly available data sets. This approach enabled us to deepen our understanding of which parts of the actual AS topology are captured in the public view and which parts remain invisible and are missing from the public view*.

A critical aspect of our search for the elusive ground truth of AS-level Internet connectivity and of the proposed pragmatic approach to constructing realistic and viable AS maps is that they both treat ASes as objects with a rich, important, and diverse internal structure, and not as generic and property-less nodes. Exploiting this structure is at the heart of our work. The nature of this AS-internal structure permeates our definition of "ground truth" of AS-level connectivity, our analysis of the available data sets, our understanding of the reasons behind and importance of the deficiencies of commonly-used AS-level Internet topologies, and our proposed efforts to construct realistic and viable maps of the Internet's AS-level ecosystem. Faithfully accounting for this internal structure can also be expected to favor the constructions of AS maps that withstand scrutiny by domain experts. Such constructions also stand a better chance to represent fully functional and economically viable AS-level topologies than models where the interconnections between different ASes are solely determined by independent coin tosses.

Acknowledgment

The authors would like to thank T. Scholl, B. Woodcock, R. Provo, J. Borkenhagen, J. Yates, S. Lee, A. Gerber, and A. Shaikh for many helpful discussions. We would also like to thank several network operators who helped us gain insight into the AS-level Internet connectivity.

References

[1] *AOL Peering Requirements*, [Online]. Available: http://www.atdn.net/settlement_free_int.shtml
[2] *AT&T Peering Requirements*, [Online]. Available: http://www.corp.att.com/peering/
[3] *CERNET BGP Feeds*, [Online]. Available: http://bgpview.6test.edu.cn/bgp-view/
[4] *European Internet Exchange Association*, [Online]. Available: http://www.euro-ix.net
[5] *Geant2 Looking Glass*, [Online]. Available: http://stats.geant2.net/lg/
[6] *Good Practices in Internet Exchange Points*, [Online]. Available: http://www.pch.net/resources/papers/ix-documentation-bcp/ix-documentation-bcp-v14en.pdf
[7] *Internet Routing Registry*, [Online]. Available: http://www.irr.net/
[8] *Packet Clearing House IXP Directory*, [Online]. Available: http://www.pch.net/ixpdir/Main.pl
[9] *PeeringDB Website*, [Online]. Available: http://www.peeringdb.com/
[10] Personal Communication with Bill Woodcock at PCH.
[11] *RIPE Routing Information Service Project*, [Online]. Available: http://www.ripe.net/

[12] *RouteViews Routing Table Archive*, [Online]. Available: http://www.routeviews.org/
[13] *Skitter AS Adjacency List*, [Online]. Available: http://www.caida.org/tools/measurement/skitter/as_adjacencies.xml
[14] *The Abilene Observatory Data Collections*, [Online]. Available: http://abilene.internet2.edu/observatory/data-collections.html
[15] H. Ballani, P. Francis, and X. Zhang, "A study of prefix hijacking and interception in the Internet," in *Proc. ACM SIGCOMM*, 2007, pp. 265–276.
[16] H. Chang, "An economic-based empirical approach to modeling the Internet inter-domain topology and traffic matrix," Ph.D. dissertation, Univ. Michigan, Ann Arbor, MI, 2006.
[17] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "Towards capturing representative AS-level Internet topologies," *Comput. Netw. J.*, vol. 44, no. 6, pp. 737–755, 2004.
[18] H. Chang, S. Jamin, and W. Willinger, "Inferring AS-level Internet topology from router-level path traces," in *SPIE ITCom*, 2001.
[19] H. Chang and W. Willinger, "Difficulties measuring the Internet's AS-level ecosystem," in *Proc. Annu. CISS*, 2006, pp. 1479–1483.
[20] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, kc claffy, and G. Riley, "AS relationships: Inference and validation," *Comput. Commun. Rev.*, vol. 37, no. 1, pp. 29–40, 2007.
[21] D. Dolev, S. Jamin, O. Mokryn, and Y. Shavitt, "Internet resiliency to attacks and failures under BGP policy routing," *Comput. Netw.*, vol. 50, no. 16, pp. 3183–3196, 2006.
[22] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 733–745, Dec. 2001.
[23] B. Hummel and S. Kosub, "Acyclic type-of-relationship problems on the internet: An experimental analysis," in *Proc. ACM IMC*, 2007, pp. 221–226.
[24] Y. Hyun, A. Broido, and kc claffy, "On third-party addresses in traceroute paths," presented at the PAM Workshop, 2003.
[25] M. Lad, R. Oliveira, B. Zhang, and L. Zhang, "Understanding the resiliency of Internet topology against false origin attacks," presented at the IEEE DSN, 2007.
[26] H. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iPlane: An information plane for distributed services," in *Proc. OSDI*, 2006, pp. 367–380.
[27] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP misconfiguration," in *Proc. ACM SIGCOMM*, 2002, pp. 3–16.
[28] Z. M. Mao, L. Qiu, J. Wang, and Y. Zhang, "On AS-level path inference," in *Proc. SIGMETRICS*, 2005, pp. 339–349.
[29] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an accurate AS-level traceroute tool," in *Proc. ACM SIGCOMM*, 2003, pp. 365–378.
[30] W. Mühlbauer, A. Feldmann, O. Maennel, M. Roughan, and S. Uhlig, "Building an AS-topology model that captures route diversity," in *Proc. ACM SIGCOMM*, 2006, pp. 195–206.
[31] W. Mühlbauer, S. Uhlig, B. Fu, M. Meulle, and O. Maennel, "In search for an appropriate granularity to model routing policies," in *Proc. ACM SIGCOMM*, 2007, pp. 145–156.
[32] R. Oliveira, B. Zhang, D. Pei, R. Izhak-Ratzin, and L. Zhang, "Quantifying path exploration in the Internet," in *Proc. ACM IMC*, Oct. 2006, pp. 269–282.
[33] R. Oliveira, B. Zhang, and L. Zhang, "Observing the evolution of Internet AS topology," in *Proc. ACM SIGCOMM*, 2007, pp. 313–324.
[34] D. Raz and R. Cohen, "The Internet dark matter: On the missing links in the AS connectivity map," in *Proc. IEEE INFOCOM*, 2006, pp. 1–12.
[35] Y. Rekhter, T. Li, and S. Hares, "Border Gateway Protocol 4," Internet Engineering Task Force, 2006, RFC 4271.
[36] M. Roughan, S. J. Tuke, and O. Maennel, "Bigfoot, sasquatch, the yeti and other missing links: What we don't know about the as graph," in *Proc. ACM IMC*, 2008, pp. 325–330.
[37] Y. Shavitt and E. Shir, "DIMES: Let the Internet measure itself," *Comput. Commun. Rev.*, vol. 35, no. 5, pp. 71–74, 2005.
[38] G. Siganos and M. Faloutsos, "Analyzing BGP policies: Methodology and tool," in *Proc. IEEE INFOCOM*, 2004, pp. 1640–1651.
[39] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *Proc. IEEE INFOCOM*, 2002, pp. 618–627.
[40] L. Subramanian, M. Caesar, C. T. Ee, M. Handley, M. Mao, S. Shenker, and I. Stoica, "HLP: A next generation inter-domain routing protocol," in *Proc. ACM SIGCOMM*, 2005, pp. 13–24.
[41] L. Wang, M. Saranu, J. M. Gottlieb, and D. Pei, "Understanding BGP session failures in a large ISP," in *Proc. IEEE INFOCOM*, 2007, pp. 348–356.
[42] J. Wu, Y. Zhang, Z. Mao, and K. Shin, "Internet routing resilience to failures: Analysis and implications," in *Proc. ACM CoNEXT*, 2007, Article No. 25.
[43] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy, "A systematic framework for unearthing the missing links: Measurements and impact," presented at the NSDI, 2007.
[44] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-level topology," *Comput. Commun. Rev.*, vol. 35, no. 1, pp. 53–61, Jan. 2005.
[45] C. Zheng, L. Ji, D. Pei, J. Wang, and P. Francis, "A light-weight distributed scheme for detecting IP prefix hijacks in real-time," in *Proc. ACM SIGCOMM*, 2007, pp. 277–288.

Ricardo Oliveira (M'08) received the B.S. degree in electrical engineering from the Engineering Faculty of Porto University (FEUP), Porto, Portugal, in 2001, and the M.S. degree in computer science from the University of California, Los Angeles (UCLA), in 2005.

He has been pursuing the Ph.D. degree in computer science at UCLA since 2005.

Mr. Oliveira is a Student Member of the Association for Computing Machinery.

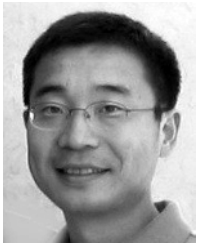
Dan Pei received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1997 and 2000, respectively, and the Ph.D. degree from the University of California, Los Angeles (UCLA), in 2005.

He is a Researcher at AT&T Research, Florham Park, NJ. His current research interests are network measurement and security.

Walter Willinger (F'05) received the diploma (Dipl. Math.) from the ETH Zurich, Zurich, Switzerland, and the M.S. and Ph.D. degrees from Cornell University, Ithaca, NY.

He is currently a Member of the Information and Software Systems Research with AT&T Labs–Research, Florham Park, NJ. He was a Member of the Technical Staff at Bellcore Applied Research from 1986 to 1996.

Dr. Willinger was a co-recipient of the 1996 IEEE W.R. G. Baker Prize Award and the 1994 IEEE W.R. Bennett Prize Paper Award. He is a Fellow of the Association for Computing Machinery and and AT&T.



Beichuan Zhang received the B.S. degree from Peking University, Beijing, China, in 1995, and the Ph.D. degree in computer science from the University of California, Los Angeles, in 2003.

He is an Assistant Professor with the Department of Computer Science, University of Arizona, Tucson. His research interests include Internet routing and topology, multicast, network measurement, and security.

Lixia Zhang (F'06) received the Ph.D. degree in computer science from the Massachusetts Institute of Technology, Cambridge.

She is currently a Professor with the Computer Science Department, University of California, Los Angeles (UCLA), which she joined in 1996. Prior to that, she was a Member of the Research Staff of Xerox Palo Alto Research Center.

Dr. Zhang is a Fellow of the Association for Computing Machinery and a Member of the Internet Architecture Board (IAB).