

Lord of the Links: A Framework for Discovering Missing Links in the Internet Topology

Yihua He, *Member, IEEE*, Georgos Siganos, *Member, IEEE*, Michalis Faloutsos, *Member, IEEE*, and Srikanth Krishnamurthy, *Senior Member, IEEE*

Abstract—The topology of the Internet at the Autonomous System (AS) level is not yet fully discovered despite significant research activity. The community still does not know how many links are missing, where these links are and finally, whether the missing links will change our conceptual model of the Internet topology. An accurate and complete model of the topology would be important for protocol design, performance evaluation and analyses. The goal of our work is to develop methodologies and tools to identify and validate such missing links between ASes. In this work, we develop several methods and identify a significant number of missing links, particularly of the peer-to-peer type. Interestingly, most of the missing AS links that we find exist as peer-to-peer links at the Internet Exchange Points (IXPs). First, in more detail, we provide a large-scale comprehensive synthesis of the available sources of information. We cross-validate and compare BGP routing tables, Internet Routing Registries, and traceroute data, while we extract significant new information from the less-studied Internet Exchange Points (IXPs). We identify 40% more edges and approximately 300% more peer-to-peer edges compared to commonly used data sets. All of these edges have been verified by either BGP tables or traceroute. Second, we identify properties of the new edges and quantify their effects on important topological properties. Given the new peer-to-peer edges, we find that for some ASes more than 50% of their paths stop going through their ISPs assuming policy-aware routing. A surprising observation is that the degree of an AS may be a poor indicator of which ASes it will peer with.

Index Terms—BGP, Internet, inter-domain, measurement, missing links, routing, topology.

I. INTRODUCTION

AN ACCURATE and complete model of the Internet topology is critical for future protocol design, performance evaluation, simulation and analysis [1]. The current initiatives of rethinking and redesigning the Internet and its operation from scratch would also benefit from such a model. However, it remains as a challenge to develop an accurate representation of the Internet topology at the AS level, despite the recent flurry of studies [2]–[9]. Currently, there is a list of

sources that contain such topological information. The list includes archives of BGP routing tables, archives of BGP routing updates, Internet Routing Registries, and archives of traceroute data. Each of these sources has its own advantages, but each of them also provides an incomplete, sometimes inaccurate view of the Internet AS topology, while these sources are often complementary. Furthermore, as far as we know, IXPs (Internet Exchange Points) have not received attention in terms of Internet topology discovery, although they play a major role in the Internet connectivity.

There are two major contributions in this work. First, we design and implement a systematic framework for discovering missing links in our current Internet topology snapshot, and provide two novelties compared to previous studies—the comprehensive synthesis of different data sources and the extraction of topological information from IXPs. Second, we apply our framework and conduct an in-depth study of the importance of these new links, and improve our understanding of the Internet topology at the AS level.

In more detail, our framework first identifies and validates a significant number of AS links by a careful cross-reference and synthesis of most known sources of information: BGP tables, traceroute, and IRR [10].¹ Second, our framework extracts significant new topological information from Internet Exchange Points (IXPs); such information is typically not used in topological studies. While prior work [11] has proposed methods to identify participating ASes at IXPs, our study greatly extends their work and overcomes certain limitations.

Note that we set a highly selective standard in our framework: we only accept edges which are verified by BGP tables or from traceroute data. In other words, we do not provide a union of the existing sources of information, but a critical synthesis. To achieve this goal, we develop a large scale traceroute-based tool, RETRO, to confirm the existence of edges, which we suspect exist.

We arrive at several interesting observations. First, we find a significant number of new edges, including 40% more edges (15%) and approximately 300% more peer-to-peer edges (65%) as compared to the widely used Oregon Routeviews data set (all available BGP routing tables, respectively). Second, most of the newly discovered edges are peer-to-peer edges: the current topological models have a bias by under-representing peer-to-peer edges. Third, most missing peer-to-peer AS links that we find are at the IXPs: Our results show that nearly 95% of the

Manuscript received June 22, 2007; revised January 13, 2008; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor A. Orda. This work was supported by the National Science Foundation under NSF NETS 22063 and NSF IDM 0208950 and a CISCO URP grant.

Y. He is with Yahoo! Inc., Sunnyvale, CA 94089 USA (e-mail: yihua@yahoo-inc.com; yhe@cs.ucr.edu).

G. Siganos is with the Telefonica Research, 08021 Barcelona, Spain (e-mail: georgos@tid.es).

M. Faloutsos and S. Krishnamurthy are with the Department of Computer Science and Engineering, University of California, Riverside, CA 92521 USA (e-mail: michalis@cs.ucr.edu; krish@cs.ucr.edu).

Digital Object Identifier 10.1109/TNET.2008.926512

¹We use IRR information as a source for optimizing the traceroute discovery effort shown later in this paper. All AS edges reported here are verified either by BGP tables or traceroute.

peer-to-peer links missed from the BGP tables are incident at IXPs. This suggests that exploring the connectivity at IXPs may help us identify hidden edges between ASes that participate at IXPs. Fourth, IRR is a good source of hints for finding new edges, especially after it is filtered using a state of the art tool [12] for this purpose.

We find that the new edges significantly change our view of the Internet AS topology, and we also identify interesting patterns of the new edges. First, the new edges change the models of Internet routing and financial implications that previous research studies may have arrived at by using the incomplete topology models.

We quantify the routing decision changes in the routing model due to the peer-to-peer edges not considered previously. We find that for some ASes (mostly of degrees 10 to 300), more than 50% of their paths stop going through a provider, compared to a less complete topology. The financial implication is that these ASes may not pay their providers to the extent that was earlier expected. Clearly, business-oriented studies should consider all peer-to-peer edges for accurate results. Second, we find that provider-customer and peer-to-peer edges have significantly different properties and they should be modeled separately: We find that the degree distribution of the provider-customer only edges can be accurately described by a power-law (with correlation coefficient higher than 99%) in all the topological instances that we examine. In contrast, degree distribution of the peer-to-peer only edges is better described by a Weibull distribution with correlation coefficient higher than 99%, which corroborates previous studies [9], [7]. Third, the degrees of the nodes of a peer-to-peer link can vary significantly: 50% of the peer-to-peer edges are between nodes whose degrees, d_1, d_2 , differ a lot either in absolute ($|d_1 - d_2| > 144$) or relative value ($\max(d_1/d_2, d_2/d_1) > 4$). This has direct implications on how we think about and model peer-to-peer edges. For instance, this observation suggests that researchers need to use caution when using the degree as an indication of whether two ASes could have a peer-to-peer relationship. Our results can provide guidelines to AS policy inference algorithms, which partly rely on the node degree. Fourth, we provide an educated guess on how many edges we may still be missing. We estimate the edges to be roughly 35% compared to the peer-to-peer edges we know at the end of this study.

This paper is an extend version of our earlier work [13], which has attracted the attention of the community. Our data set² has been downloaded by more than 50 different universities and research institutes since January 2007. In this version, we provide more details about our data, and discuss in more detail our results and our methodology of inferring IXP participants.

The rest of this paper is organized as follows. We review the data sources and previous work in Section II. In Section III, we present our framework and the motivation behind its design. In Section IV, we quantify the impact of our new found AS links. We introduce our methods to identify the IXP participants in Section V. In Section VI, we summarize our work.

II. BACKGROUND

A. Data Sources and Their Limitations

In this section, we describe the most popular data sources and their two main limitations: incompleteness and a bias in the nature of the discovered links.

BGP routing table dumps are probably the most widely used resource that provides information on the AS Internet topology. Each table entry contains an AS path, which corresponds to a set of AS edges. Several sites collect tables from multiple BGP routers, such as Routeview [14] and RIPE/RIS [15]. An advantage of the BGP routing tables is that their link information is considered reliable. If an AS link appears in a BGP routing table dump, it is almost certain that the link exists. However, limited number of vantage points makes it hard to discover a more complete view of the AS-level topology. A single BGP routing table has the union of “shortest” or, more accurately, preferred paths with respect to this point of observation. As a result, such a collection will not see edges that are not on any preferred path for this point of observation. Several theoretical and experimental efforts explore the limitations of such measurements [16], [17]. Worse, such incompleteness may be statistically biased based on the type³ of the links. Some types of AS links are more likely to be missing from BGP routing table dumps than other types. Specifically, peer-to-peer links are likely to be missing due to the selective exporting rules of BGP. Typically, *a peer-to-peer link can only be seen in a BGP routing table of these two peering ASes or their customers*. A recent work [9] discusses in depth this limitation.

BGP updates are used in previous studies [3], [5] as a source of topological information and they show that by collecting BGP updates over a period of time, more AS links are visible. This is because as the topology changes, BGP updates provide transient and ephemeral route information. However, if the window of observation is long, an advertised link may cease to exist [3] by the time that we construct a topology snapshot. In other words, BGP updates may provide a superimposition of a number of different snapshots that existed at some point in time. Recently, Oliveira *et al.* [18] explicitly distinguished this commonly overlooked “liveness problem” from the “completeness problem”, which is the central topic of this paper. Note that BGP updates are collected at the same vantage points as the BGP tables in most collection sites. Naturally, topologies derived from BGP updates share the same statistical bias per link type as from BGP routing tables: peer-to-peer links are only to be advertised to the peering ASes and their customers. This further limits the additional information that BGP updates can provide currently. On the other hand, BGP updates could be useful in revealing ephemeral backup links over long period of observation, along with erroneous BGP updates.

By using traceroute, one can explore IP paths and then translate the IP addresses to AS numbers, thus obtaining AS paths. Similar to BGP tables, the traceroute path information is considered reliable, since it represents the path that the packets actu-

²<http://www.cs.ucr.edu/~yhe/LordOfLinks/>

³Most ASes peer with each other with two types of links: the provider-customer links and peer-to-peer links. Normally, customer ASes pay their providers for traffic transit, and ASes with peer-to-peer relationship exchange traffic with no or little cost to each other.

ally traverse. On the other hand, a traceroute server explores the routing paths from its location towards the rest of the world, and thus, the collected data has the same limitations as BGP data in terms of completeness and link bias. One additional challenge with the traceroute data is the mapping of an IP path to an AS path. The problem is far from trivial, and it has been the focus of several recent efforts [19]–[21].

Internet Routing Registry (IRR) [10] is the union of a growing number of world-wide routing policy databases that use the Routing Policy Specification Language (RPSL). In principle, each AS should register routes to all its neighbors (that reflect the AS links between the AS and its neighbors) with this registry. IRR information is manually maintained and there is no stringent requirement for updating it. Therefore, without any processing, AS links derived from IRR are prone to human errors, could be outdated or incomplete. However, the up-to-date IRR entries provide a wealth of information that could not be obtained from any other source. A recent effort [12] shows that, with careful processing of the data, one can extract a nontrivial amount of correct and useful information.

B. Related Work and Comparison

There has been a large number of measurements studies related to topology discovery, with different goals, at different times, and using different sources of information.

Our work has the following characteristics that distinguish it from most previous other efforts, such as [9], [2]: 1) We make extensive use of topological information from the Internet Exchange Points to identify more edges. It turns out that IXPs “conceal” many links which did not appear in most previous topology studies. 2) We use a more sophisticated, comprehensive and thorough tool [12] to filter the less accurate IRR data, which was not used by previous studies. 3) We employ a “guess-and-verify” approach for finding more edges by identifying potential edges and validating them through targeted traceroutes. This greatly reduced the number of traceroutes that were needed. 4) We accept new edges conservatively and only when they are confirmed by a BGP table or a traceroute. In contrast, some of the previous studies included edges from IRR without confirming them with traceroute.

The most relevant previous work is done by Chang *et al.* [2] with data collected in 2001. They identify new edges by looking at several sources of topological information including BGP tables and IRR. They estimate that 25%-50% AS links were missing from Oregon Routeview BGP table, the most commonly used data set for AS topology studies. Their work was an excellent first step towards a more complete topology.

In a parallel effort, Cohen and Raz [9] identify missing links in the Internet topology. Our studies corroborate some of the observations there. Note that, their work does not include an exhaustive measurement, data collection and comparison effort as our work. For example, IXP information was not used in their work.

Several other interesting measurement studies exist. NetDimes [4] is an effort to collect large volumes of host-based traceroute information. The key here is to increase the number of traceroute points by turning cooperative end hosts into observation points. The challenge now becomes the measurement

noise removal, the collection, and processing of the information [22]. Our approach and NetDimes could complement and leverage each other towards a more complete and accurate topology. Donnet *et al.* [23] propose efficient algorithms for large-scale topology discovery by traceroute probes. Rocket-fuel [24] explores ISP topologies using traceroutes. In [5], the authors examine the information contained in BGP updates.

Most of these studies and our work seek a complete snapshot of the Internet topology. In other words, short-lived *backup* links are most likely not included in most such studies. Some ASes have such links, which normally are not “visible” unless the primary links are down. Recently, active BGP probing [8] has been proposed as a method for identifying backup AS links, and this could complement our work and the efforts mentioned above.

There are several efforts that study the topology and they would benefit from an accurate and complete topology. A plethora of efforts attempts to model the topology and to generate realistic topologies (e.g., [25]). Some studies [16], [26] document the limitations of the sources of topological information, but without necessarily attempting to identify a more complete topology. A recent study [7] models the evolution of the Internet topology by investigating the process of AS peerings. Another recent work [18] models the evolution using a constant rate birth–death process. Our work can be seen as a basis that can provide more complete and accurate information for such studies.

The exhaustive identification of *IXP participants* has received limited attention. Most previous work focuses on identifying the existence of IXPs. Xu *et al.* [11] develop what appears to be the first systematic method for identifying IXP participants. Inspired by their work, our approach subsumes their method, and thus, it provides more complete and accurate results (see Section V).

III. FRAMEWORK FOR FINDING MISSING LINKS

In this section, we present a systematic framework for extracting and synthesizing the AS level topology information from different sources. The different sources have complementary information of variable accuracy. Thus, we cannot just simply take the union of all the edges. A careful synthesis and cross-validation is required. At the same time, we are interested in identifying the properties of the missing AS links.

In a nutshell, our study arrives at three major observations regarding the properties of the missing AS links: 1) most of the missing AS edges are of the peer-to-peer type; 2) most of the missing AS edges from BGP tables appear in IRR; and 3) most newfound AS edges are incident at IXPs. At different stages of the research, these three observations direct us to discover even more edges, some of which do not appear in any other source of information currently.

We present an overview of our work in order to provide the motivation for the different steps that we take. We start with the data set from Oregon routeviews BGP table Dump (*OBD*) [14], the BGP table dumps collected at route-views.oregon-ix.net, which is by far the most widely used data archive. Our work consists of four main steps.

TABLE I
TOPOLOGICAL DATA SETS USED IN OUR STUDY

OBD	The Oregon routeviews BGP table dump
BD	OBD and other additional BGP table dumps
IRRnc	IRR edges processed by Nemecis with non-conflicting policy declarations
IRRdual	IRRnc edges correctly declared by both adjacent ASes
BD+IRR	BD and the edges of IRRdual confirmed by RETRO
IXPall	Union of cliques of IXP participants
ALL	BD+IRR and the potential IXP edges that are confirmed by RETRO

TABLE II
STATISTICS OF THE TOPOLOGIES

Name	Nodes	Edges	p-c	p-p
<i>OBD</i>	19.8k	42.6k	36.7k	5.5k
<i>BD</i>	19.9k	51.3k	38.2k	12.7k
<i>BD+IRR</i>	19.9k	56.9k	38.2k	18.3k
<i>ALL</i>	19.9k	59.5k	38.2k	20.9k

A. BGP routing tables: We consider the AS edges derived from multiple BGP routing table dumps [3], and compare them to the Routeview data (OBD). The question we try to answer is what is the information that the new BGP tables bring. We use the term *BD* to refer to the union data from all available BGP table Dumps. Table I lists the acronyms for our data sets.

B. IRR data: We systematically analyze the IRR [10] data and identify topological information that seems trustworthy by Nemecis [12]. We follow a conservative approach, given that IRR may contain some outdated and/or erroneous information. We do not accept new edges from IRR, even after our first processing, unless they are confirmed by traceroutes (using our RETRO tool). Overall, we find that IRR is a good source of hints for missing links. For example, we discover that more than 80% of the new edges found in the new tables (i.e., the AS edges in BD but not in OBD) already exist in IRR. Even compared to BD, IRR has significantly more edges, which are validated by RETRO as we explain below.

C. IXPs and potential edges: We identify a set of potential IXP edges by applying our methodology on inferring IXP participants from Section V. We find that many of the peer-to-peer edges missing from the different data sets could be IXP edges.

D. Validation using RETRO: We use our traceroute tool, RETRO, to verify potential edges from IRR and IXPs. First, we confirm the existence of many potential edges we identified in the previous steps. We find that more than 94% of the RETRO-verified AS edges in IRR indeed go through IXPs. We also discover edges that were not previously seen in either the BGP table dumps or IRR. In total, we have validated 300% more peer-to-peer links than those in the OBD data set.

The statistics of the topologies generated from the different data sets in our study are listed in Table II.

A. The New Edges From a BGP Table Dump

We collect multiple BGP routing table dumps from various locations in the world, and compare them with OBD. On May 12, 2005, we collected 34 BGP routing table dumps from the Oregon route collectors [14], the RIPE/RIS route collectors [15] and public route servers. Several other route collectors were not

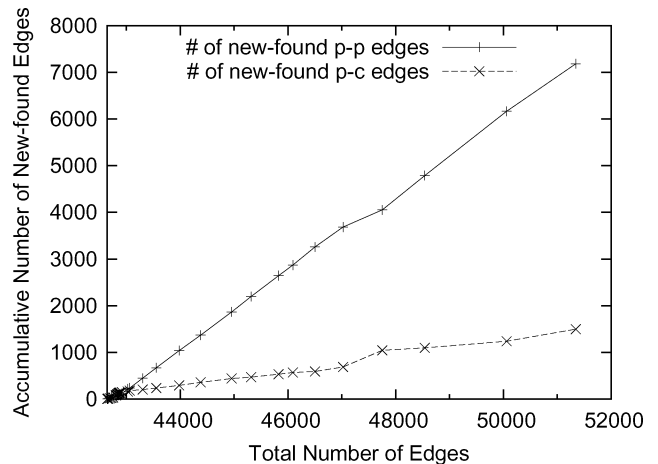


Fig. 1. Most new edges in BD but not in OBD are peer-to-peer edges.

operational at the time that the data was collected and therefore, we do not include them in this study. For each BGP routing table dump, we extract its “AS_PATH” field and generate an AS topology graph. We then merge these 34 graphs into a single graph and delete duplicate AS edges if any. The resulting graph, which is named as *BD* (*BGP Dumps*), has 19 950 ASes and 51 345 edges. The statistics of *BD* are similar to what was reported in [3]. Interestingly, *BD* has only 0.5% additional ASes, but 20.4% more AS edges as compared with *OBD*.

To study the business relationships of these edges, we use the PTE algorithm [27], which seems to outperform most previous such approaches. Specifically, it significantly increases the accuracy (over 90%) of inferring peer-to-peer AS links. Most of the AS edges are classified into three basic types on the basis of business relationships: provider-customer, peer-to-peer and sibling-to-sibling. Among them, sibling-to-sibling links only account for a very small (0.12%) portion of the total AS edges and we do not consider them in this study. We count the number of peer-to-peer (or “p-p” for short) and provider-customer (or “p-c” for short) AS links for each BGP routing table. The statistics for dumps with significant number of new edges are shown in Table III.

For comparison purposes, we pick the most widely used AS graph *OBD* as our baseline graph. For each of the other BGP routing tables, we examine the number of additional AS edges that do not appear in *OBD*, as classified by their business relationship. As shown in Table III, from each of the BGP routing tables that provides a significant number of new edges to *OBD*, most of the newfound edges are of the peer-to-peer type.

BGP table biases: underestimating the peer-to-peer edges. A closer look at the data reveals an interesting dichotomy: 1) most edges in a BGP table are provider-customer; and 2) given a set of BGP tables, most new edges in an additional BGP table are peer-to-peer type. We can see this by plotting the types of new edges as we add the new tables. In Fig. 1, we plot the cumulative number of new found peer-to-peer edges and provider-customer edges versus the total number of edges. To generate this plot, we start with *OBD* with 42 643 AS edges and merge new AS edges derived from the BGP table dumps other than *OBD*, one table dump at a time, sorted by the number of

TABLE III
COLLECTION OF BGP TABLE DUMPS (AS OF MAY 12, 2005)

Route collector or Router server name	Location or AS Num	# of Nodes	# of Edges	# of edges with type inferred			edges not in OBD	edges not in OBD w/ type		
				total	p-p	p-c		total	p-p	p-c
Oregon Routeview route collectors										
route-views.oregon-ix.net(<i>OBD</i>)	Oregon, US	19843	42643	42570	5551	36766	0	0	0	0
route-views2.oregon-ix.net	Oregon, US	19837	41274	41230	4464	36514	1029	1028	835	191
route-views.eqix.routeviews.org	Virginia, US	19650	34889	34876	1027	33640	674	674	530	143
route-views.linx.routeviews.org	London, UK	19655	37259	37246	3246	33765	2511	2511	2188	319
route-views.isc.routeviews.org	Palo Alto, US	19753	36152	36139	1915	34004	784	783	663	118
route-views.wide.routeviews.org	Tokyo, JP	19649	26974	26963	636	26183	35	35	26	8
RIPE RIS route collectors										
rrc00.ripe.net	Amsterdam, NL	19770	36479	36465	1641	34605	655	654	543	111
rrc01.ripe.net	London, UK	19640	34193	34180	1121	32855	617	617	512	105
rrc02.ripe.net	Paris, FR	3966	4260	4256	27	4204	3	3	0	3
rrc03.ripe.net	Amsterdam, NL	19737	39147	39129	3850	35042	3233	3228	2609	616
rrc04.ripe.net	Geneva, CH	19694	29840	29827	660	28997	452	451	366	85
rrc05.ripe.net	Vienna, AT	19765	32676	32659	1122	31324	1095	1091	658	432
rrc07.ripe.net	Stockholm, SE	19618	32811	31797	1219	30394	804	803	724	79
rrc10.ripe.net	Milan, IT	19578	29225	29213	342	28704	122	122	102	20
rrc11.ripe.net	New York, US	19592	29784	29772	716	28892	441	441	377	64
rrc12.ripe.net	Frankfurt, DE	19628	33841	33827	2024	31606	1611	1610	1417	193
Public route servers										
route-server.belwue.de	AS553	19637	25208	25195	194	24836	93	92	21	69
route-views.on.bb.telus.com	AS852	19551	24893	24883	129	24610	30	30	24	5
route-views.ab.bb.telus.com	AS852	19551	24894	24884	129	24610	30	30	24	5
route-sever.ip.tiscali.net	AS3257	19652	24808	24798	168	24498	51	51	1	48
route-server.gblx.net	AS3549	19263	24878	24866	57	24680	54	54	2	51
route-server.eu.gblx.net	AS3549	19267	26066	26054	60	25859	55	55	2	52
route-server.savvis.net	AS3561	19210	24630	24620	31	24464	3	3	1	1
route-server.as5388.net	AS5388	19526	25049	25039	133	24785	58	58	2	55
route-server.gt.ca	AS6539	19472	25822	25981	187	25477	62	62	1	60
route-server.as6667.net	AS6667	19459	25073	25063	179	24754	83	83	15	67
route-server.sunrise.ch	AS6730	15414	19637	19627	198	19307	32	32	20	11
route-server.he.net	AS6939	19396	25661	25650	207	25314	70	70	9	60
route-server.ip.att.net	AS7018	19572	26822	26809	83	26581	67	67	0	66
route-views.optus.net.au	AS7474	19606	30045	30032	507	29363	338	337	223	113
route-server.wcg.net	AS7911	19420	25566	25556	89	25325	63	63	0	62
route-server.colt.net	AS8220	19261	27988	27975	237	27580	128	128	30	180
route-views.bmcag.net	AS9132	19461	24639	24628	128	24362	17	17	8	8
route-server.rhein-main-saar.net	AS15837	19540	26003	25991	492	25359	312	311	223	87
Union (<i>BD</i>)		19950	51345	51259	12734	38265	8702	8689	7183	1499

new edges they provide. At the end, when all the BGP table dumps in our data set are included, we obtain the graph *BD*; this has 51 345 AS edges in total. Among these edges, there are 7183 peer-to-peer edges and 1499 provider-customer edges that do not exist in the baseline graph *OBD*. Clearly, Fig. 1 demonstrates that we discover more peer-to-peer AS edges than provider-customer edges when we increase the number of vantage points. Furthermore, the ratio of the number of new found peer-to-peer edges to the number of new found provider-customer edges is almost constant given that the two curves (corresponding to the new found p-p edges and the p-c edges) in Fig. 1 are almost straight lines.

The percentage of peer-to-peer edges increases with the number of BGP tables. A complementary observation is that for a BGP-table-based graph, the more complete it is (in number of edges), the higher the percentage of peer-to-peer links. For example, the AS graph derived from rrc12.ripe.net has 33 841 AS edges, 2024 (5.98%) of which are peer-to-peer edges. On the other hand, the more complete AS graph *OBD* has 42 643 edges, and 5551 (13.0%) of these edges are peer-to-peer edges.

The union graph *BD* has an even higher percentage (24.8%) of peer-to-peer links.

The above observations strongly suggest that in order to obtain a more complete Internet topology, one should pay more attention to discovering peer-to-peer links.

B. Exploring IRR

We carefully process the IRR information to identify potential new edges. Recall that we do not add any edges until we verify them with RETRO later in this section.

We extract AS links from IRR on May 12, 2005 and classify their business relationships using Nemecis [12] as per the exporting policies of registered ISPs. The purpose of using Nemecis to filter the IRR is that, Nemecis can successfully eliminate most badly defined or inconsistent edges and, it can infer with fair accuracy the business relationships of the edges.

There are 96 654 AS links in total and they are classified into three basic types in terms of their relationships: peer-to-peer, customer-provider and sibling-to-sibling. Sometimes two ASes register conflicting policies with each other. For example, AS_A

TABLE IV
AS EDGES IN IRR (MAY 12, 2005) WITHOUT RELATIONSHIP CONFLICT

Name of Graphs	# of non-0 degree Nodes	# of Edges	Avg Degree	Perc. of total IRR edges	Perc. of IRR edges without conflict
<i>IRRnc</i>	16952	89540	10.56	92.6%	100.0%
<i>peerIRRnc</i>	6619	49411	14.93	51.1%	55.2%
<i>pcIRRnc</i>	15277	37619	4.925	38.9%	42.0%
<i>siblingIRRnc</i>	2277	2510	2.204	2.6%	2.8%
<i>peerIRRdual</i>	1561	18453	23.64	19.1%	20.6%
<i>pcIRRdual</i>	6298	8748	2.778	9.1%	9.8%
<i>siblingIRRdual</i>	226	143	1.265	0.1%	0.1%

may register AS_B as a customer while AS_B registers AS_A as a peer. There are 7114 or 7.4% of such AS links and we exclude them in our data analysis. We call the remaining edges *nonconflicting IRR edges* or *IRRnc*. Considering the different types of policies, this set can be decomposed into three self-explanatory sets: *pcIRRnc*, *peerIRRnc* and *siblingIRRnc*. From these edges, we define the set *IRRdual* to include the edges for which both adjacent ASes register matching relationships. (Contrarily, *IRRnc* includes edges for which only one AS registers a peering relationship while the other AS does not register at all.) Similarly, the *IRRdual* set can be decomposed by type of edge into three sets: *pcIRRdual*, *peerIRRdual* and *siblingIRRdual*.

The statistics of these data sets are summarized in Table IV. We notice that the number of edges in the more reliably defined *IRRdual* set is significantly less than that of the *IRRnc*. In other words, AS edges in *IRRdual* and its subsets (*peerIRRdual*, *pcIRRdual* and *siblingIRRdual*) are fewer but we are more confident about: a) their existence, and b) their business relationships.

We make the following two observations.

1) IRR is a good source of hints for missing edges. We perform the following thought experiment: *knowing only the OBD data set, would IRR be a good source of potential edges?* We compare the edges in graph *BD* but not in graph *OBD* with the edges in IRR. We find that 83.3% of these edges exist in IRR: 7251 from a total of 8702 new edges. This high percentage suggests that the IRR can potentially be a source for finding new edges. We also notice that from among these 7251 edges, 6302 are classified in terms of their business relationships by Nemezis [12]. From among these classified edges, 5303 edges are of the peer-to-peer type and only 832 are of the provider-customer type. This confirms the result shown in Fig. 1, where most new found AS edges are of the peer-to-peer type. Recall that, for Fig. 1, the business relationships are inferred by the PTE algorithm [27], instead of Nemezis [12], which we use here. Both algorithms give quantitatively similar results which provides high credibility to both the data and the interpretations.

2) IRR has many more edges compared to our most complete BGP-table graph (*BD*). Motivated by the observation above, we examine the number of AS edges in IRR that are not included in *BD*. Table V summarizes the number and the type of IRR AS edges that do not appear in *BD*. From among the IRR AS edges inferred as nonconflicting types, 71.1% are missing from *BD*. The percentage is especially high for peer-to-peer edges: 80.7% of the peer-to-peer AS edges in IRR are missing from *BD*. This suggests that there may be many IRR links that exist but are yet to be verified. We also notice that 59.7% of the

TABLE V
PERCENTAGE OF IRR EDGES MISSING FROM *BD*

Name	# of edges	# of edges NOT in <i>BD</i>	# of edges Missing Perc.
<i>IRRnc</i>	89,540	63,660	71.1%
<i>peerIRRnc</i>	49,411	39,894	80.7%
<i>pcIRRnc</i>	37,619	22,466	59.7%
<i>siblingIRRnc</i>	2,510	1,300	51.8%

provider-customer AS edges are missing. At this point, we can only speculate that most of these missing provider-customer AS edges represent backup links.

C. IXPs and Missing Links

Note that, when two ASes are participants at the same IXP, it does not necessarily mean that there is an AS edge between them. If two participating ASes agree to exchange traffic through an IXP, this constitutes an AS edge, which we call an *IXP edge*. Many IXP edges are of peer-to-peer type, although customer-provider edges are also established.

Identifying IXP edges requires two steps: 1) we need to find the IXP participants, and 2) we need to identify which edges exist between the participants. We defer a discussion of our method and tool on how to find the IXP participants to Section V. However, even when we know the IXP participants, identifying the edges is still a challenge: not all participants connect with each other. In addition, the peering agreements among the IXP participants are not publicly known.

We start with a superset of the real IXP edges that contains all possible IXP edges: we initially assume that the participants of each IXP form a clique. We denote by *IXPall* the set of all edges that make up all of these cliques. *IXPall* contains 141 865 distinct AS edges.

Potential missing edges and IXP edges. We revisit the previous sets of edges we have identified and check to see if they could be IXP edges. First, we look at the peer-to-peer AS edges that appear in *BD* but not in *OBD*. We call this set of AS edges *peerBD-OBD*. Here we use the minus sign to denote the difference between two sets: *A-B* is the set of entities in set *A* but not in set *B*. Second, we look at the AS edges that appear in *peerIRRnc* but not in the graph *BD*. We call this set of links *peerIRRnc-BD*. These AS links are the ones that are potentially missing from *BD*. We define the *peerIRRdual* links not in *BD* as *peerIRRdual-BD*.

Having made this classification, we compare each class with the super set, *IXPall*, of edges that we constructed earlier. The statistics are shown in Table VI. With our first comparison, we find that approximately 86% of the edges in *peerBD-OBD* are in *IXPall* and hence, are potentially IXP edges. Next, we observe that 60% of the edges in *peerIRRnc-BD* and 83% of the edges in *peerIRRdual-BD* are in *IXPall*. Thus, if they exist, they could be IXP edges.

In summary, the analysis here seems to suggest that, most of the peer-to-peer AS links missing from the BGP dumps but present in IRR are potentially IXP edges.

D. Validating Links With RETRO

With the work so far, we have identified sets of edges and obtained hints on where to look for new edges: 1) most missing

TABLE VI
MANY MISSING PEER-TO-PEER LINKS ARE AT IXPs

Name	# of Edges	\cap <i>IXPall</i>	Perc.
<i>peerBD-OBD</i>	7183	6197	86%
<i>peerIRRnc-BD</i>	39894	23979	60%
<i>peerIRRdual-BD</i>	13905	11477	83%
<i>BD-OBD</i>	8702	6910	79%

links are expected to be the peer-to-peer type; 2) IRR seems to be a good source of information; 3) many missing edges are expected to be IXP edges.

However, as we have noted before, the peer-to-peer edges learned through the IRRs and *IXPall* are not guaranteed to exist. Therefore, in this section we focus on validating their existence to the extent possible. *Note here that with the validation, we eliminate stale information that may still be present in the IRR and IXP data sources.* To verify the existence of the edges in *peerIRRnc-BD*, we would like to witness these edges on traceroute paths. Typically, when a traceroute probe passes through an IXP edge between AS A and AS B, it will contain the following sequence of IP addresses: $[IP_{AS_A}, IP_{IXP}, IP_{AS_B}]$. If such a pattern is observed with our traceroute probes, it is almost certain that an IXP edge between AS A and AS B exists.

We first tried to use the Skitter⁴ traces as our verification source; however, we soon found that it was not suitable for our purposes. Between May 8 and May 12 in 2005, we collected a full cycle of traces from each of the active Skitter monitors. Despite a total number of 21 363 562 individual traceroute probes in the data set, we were only able to confirm 399 IXP edges in *peerIRRnc-BD*. The reason could be that the monitors were not in the “right” place to discover these edges: the monitors should be at the AS adjacent to that edge, or at one of the customers of those two ASes. With the limited number of monitors (approximately two dozen active ones) in Skitter, it is difficult to witness and validate many of the peer-to-peer AS edges.

To address this limitation, we develop a tool for detecting and verifying AS edges. We employ public traceroute servers (e.g., [28]) to construct RETRO (REverse TraceRoute), a tool that collects traceroute server configurations, send out traceroute requests, and collect traceroute results dynamically. Currently, we have a total of 404 reverse traceroute servers which contain more than 1200 distinct and working vantage points. These vantage points cover 348 different ASes and 55 different countries. We will see later that RETRO is very efficient in discovering missing peer-to-peer edges—for the dataset *peerIRRnc-BD*, we are able to confirm 5646 edges from less than 10 000 traceroutes.

With the RETRO tool, we conduct the following procedure to verify AS edges in the *peerIRRnc-BD* set. For each edge in *peerIRRnc-BD*, we find out if there are any RETRO monitors in at least one of the two ASes incident on the edge. For about 2/3 of the edges in *peerIRRnc-BD*, we do not have a monitor in either of the two ASes on the edge. If there is at least one monitor, we try to traceroute from that monitor to an IP that belongs to the other AS on the edge. There are two problems in finding the right IP address to traceroute to. First, some ASes do not announce or can not be associated with any IP prefixes and thus, we are not able to traceroute to these ASes. Second, most of

TABLE VII
RETRO VERIFIES PEER-TO-PEER LINKS IN IRR MISSING FROM BD

Name	# of edges	# of RETRO candidates	# of confirmed peering		
			total	via IXP	direct
<i>peerIRRnc-BD</i>	39894	8791	5646	5317	329
<i>peerIRRdual-BD</i>	13905	4487	3529	3351	178

the rest of the ASes announce a large range (equal to or more than 256, i.e., a full/24 block) of IP addresses. To maximize our chances of performing a successful traceroute, we choose a destination from the list of IP addresses that has been shown to be reachable by at least one of the Skitter monitors. We then trigger RETRO to generate a traceroute from the selected monitor to the destination IP address that we choose. We call this set of traceroutes *RETRO_TRACE1*.

Most newfound peer-to-peer links are incident at IXPs. We define a *candidate* to be a potential edge between two ASes, which satisfy the following two conditions: 1) we have a RETRO monitor located in one of the two ASes, and 2) there is at least one IP address from the other AS reachable by the traceroute probe performed from the RETRO monitor. We have 8791 such “candidates” for the potential AS edges in *peerIRRnc-BD*. By appropriately performing traceroutes on candidates, we get traceroute paths. In these paths, we search for two patterns for each candidate (AS_A, AS_B): a) $[IP_{AS_A}, IP_{AS_B}]$, and b) $[IP_{AS_A}, IP_{IXP}, IP_{AS_B}]$. If either of the two patterns appears, it is almost certain that the AS edge between AS_A and AS_B exists either as a) a direct edge or b) as an IXP edge, respectively. The results that we obtain at the end of the above process are summarized in Table VII.

Among 8791 candidates in *peerIRRnc-BD*, RETRO is able to confirm that a total of 5646 edges indeed exist. The existence of the rest of the candidates does not show in our RETRO data. Note that this method can only confirm the presence, but not prove the absence of an edge. It could very well be that the traceroute does not pass through the right path. An interesting observation is that 94.2% (5317/5646) of the new edges are IXP edges. This could always be an artifact introduced by biases from the measurement approach. Another explanation could be that the peer-to-peer links between middle or low ranked ASes (national or regional ISPs) are typically underrepresented in BGP tables. For those ASes, peering with other ASes at IXPs is a much more cost-efficient way than by building private peering links one by one. Our result strongly suggests that in order to look for missing peer-to-peer links from BGP tables, we should examine IXPs more carefully.

Discover edges not observed in BGP tables or IRRs. From the results so far, we suspect that the missing edges are often IXP edges. Following this pattern, we identify and confirm edges that previously had not been observed in any other data source.

We consider those AS edges in *IXPall* that are neither in *BD* nor in *IRRnc*, and call them *IXPall-BD-IRR*. We then attempt to trace these edges by using RETRO. We call this set of traceroute *RETRO_TRACE2*. The results from our experiments are summarized in Table VIII.

We find 2603 new AS edges from 17 640 RETRO candidate paths. The percentage of confirmed new AS edges is 14.8%. This is much lower than what we see with *peerIRRnc-BD*. This

⁴<http://www.caida.org/tools/measurement/skitter/>

TABLE VIII
RETRO VERIFIES AS EDGES NOT IN *BD* AND *IRRnc*

Name	# of edges	# of RETRO candidates	# of confirmed peering		
			total	via IXP	direct
<i>IXPall-BD-IRR</i>	100,076	17,640	2,603	2,407	196

is due to the fact that *IXPall* is an overly aggressive estimate. In addition, we have already identified that many edges from *IXPall* are in the previous sets (*BD* and *peerIRRnc-BD*).

We also notice that there is a small number of confirmed edges that are shown to exhibit direct peering instead of peering at some IXP. A closer look reveals that many of such cases are due to the fact that a small number of routers do not respond with ICMP messages with the incoming interfaces, and therefore, the IXP IP address, which is supposed to be returned by the traceroute, is “skipped.” Note that this phenomenon does not stop us from identifying the edge. It just makes us underestimate the percentage of IXP edges among the confirmed edges.

IV. SIGNIFICANCE OF THE NEW EDGES

In this section, we identify properties of the new edges. Then, we examine the impact of the new edges on the topological properties of the Internet. Finally, we attempt to extrapolate and estimate how many edges we may still be missing. Note that we quantify the impact of the new edges that we were able to find in the previous section. Clearly, any bias in the discovery process will affect the observations in this section.

A. Patterns of the Peer-To-Peer Edges

We study the properties exhibited by nodes that peer. Therefore, we examine the degrees, d_1 and d_2 , of the two peering nodes that make up each peer-to-peer edge. Let us clarify that the degrees d_1 and d_2 include both peer-to-peer and provider-customer edges. One would expect that d_1 and d_2 would be “comparable.” Intuitively, one would expect that the degree of an AS is *loosely* related to its importance and its place in the AS hierarchy; we expect ASes to peer with ASes at the same level.

However, we find that *the node degree of the nodes connected with a peer-to-peer link can differ significantly*. We compare the two degrees using their ratio and absolute difference. Note that these two metrics provide complementary view of difference, which leads to the following two findings. 1) Close to 78% of the peer-to-peer edges connect ASes whose degrees differ by a factor of 2 or more. In Fig. 2(a), we plot the CDF of the distribution of the ratio $\min(d_1, d_2)/\max(d_1, d_2)$ of the peer-to-peer edges. Another observation is that 45% of the peer-to-peer edges connect nodes whose degrees differ by a factor of 5 or more. This is a surprisingly large difference. One might argue that this is an artifact of having peer-to-peer edges between low degree nodes, say $d_1 = 2$ and $d_2 = 11$, whose absolute degree difference is arguably small. This is why we examine the absolute difference of the degrees next. 2) 35% of the peer-to-peer edges have nodes with an absolute difference greater than 215. In Fig. 2(b), we plot the CDF of the distribution of the absolute value $|d_1 - d_2|$, where d_1 and d_2 remain as defined earlier. Another interesting observation is that approximately half of the peer-to-peer edges have a degree difference larger than 144. Differences of 144 and 215 are fairly large if we consider

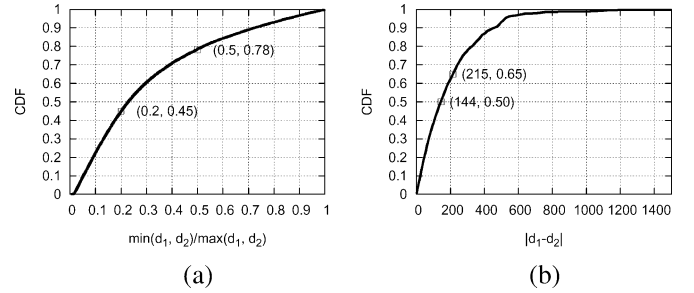


Fig. 2. (a) Degree ratio distribution and (b) degree difference distribution of all peer-to-peer AS links in the Internet.

that roughly 70% of the nodes have a degree less than 4. We intend to investigate why quite a few high degree ASes establish peer relationship with low degree ASes in the future.

B. Impact on the Internet Topology

We study the effect of the newfound peer-to-peer edges on some commonly used Internet properties. Among all the properties that we examined, we show the ones that lead to the most interesting observations.

1) *The Degree Distribution*: There has been a long debate on whether the degree distribution of the Internet at the AS level follows a power-law [29]–[31], [2]. This debate is partly due to the absence of a definitive statistical test. For example, in Fig. 3 top left, we plot the complementary cumulative distribution functions (CCDF), on a log-log scale, of the graph *ALL* defined earlier in Table I. The distribution is highly skewed, and the correlation coefficient of a least square errors fitting is 98.9%. However, one could still use different statistical metrics and argue against the accuracy of the approximation [31].

Furthermore, the answer could vary depending on which source we think is more complete and accurate, and the purpose or the required level of statistical confidence of a study. For example, if we go with *IRRdual*, which is a subset of the AS edges recorded in IRR filtered by Nemecis, the correlation coefficient is only 93.5%; see Fig. 3 top right.

To settle the debate, we propose a reconciliatory divide-and-conquer approach. We propose to model separately the degree distribution according to the type of the edges: provider-customer and peer-to-peer. We argue that this would be a more constructive approach for modeling purposes. This decomposition seems to echo the distinct properties of the two edge types, as discussed in a recent study of the evolution on the Internet topology [7].

In Fig. 3, we show an indicative set of degree distribution plots for graph *ALL* on the left column and *IRRdual* on the right. We show the distributions for the whole graph (top row), the provide-customer edges only (middle row), and the peer-to-peer edges only (bottom row). We display the power-law approximation in the first two rows of plots and the Weibull approximation in the bottom row of plots.

We observe the following two properties. 1) The provider-customer-only degree distribution can be accurately approximated by a power-law. The correlation coefficient is 99.5% or higher in the plots of Fig. 3 in the middle row. Note that, although the combined degree distribution of *IRRdual* does not

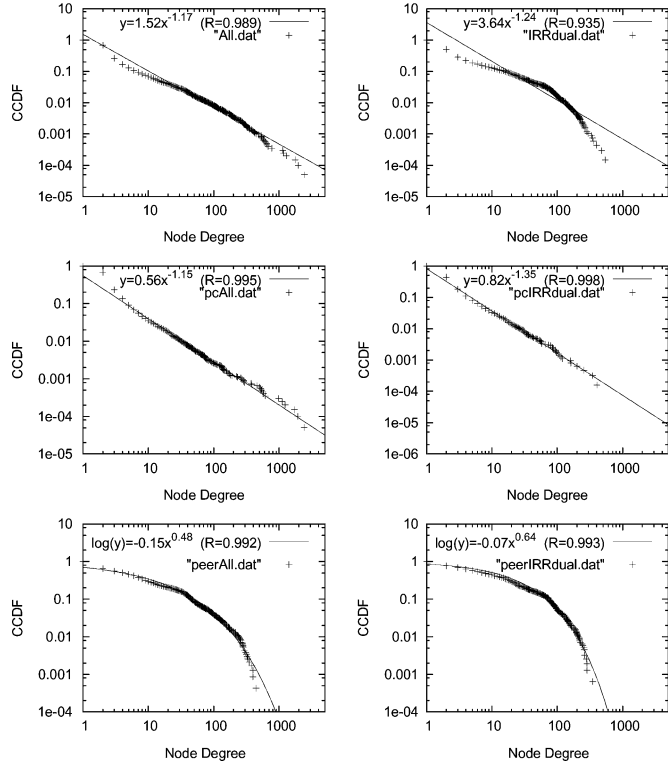


Fig. 3. The degree distributions of *ALL* (left) and *IRRDual* (right) in the top row, their provider-customer degree distributions in the middle row, and their peer-to-peer degree distributions in the bottom row.

follow a power law (top row right), its provider-customer subgraph follows a strict power law (middle row right). 2) The peer-to-peer-only degree distribution can be accurately approximated by a Weibull distribution. The correlation coefficient is 99.2% or higher in the plots of Fig. 3 in the bottom row.

It is natural to ask why the two distributions differ. We suggest the following explanation. Power-laws are related to the rich-get-richer behavior: low degree nodes “want” to connect to high degree nodes. For provider-customer edges, this makes sense: an AS wants to connect to a high-degree provider, since that provider would likely provide shorter paths to other ASes. This is less obviously true for peer-to-peer edges. If AS1 becomes a peer of AS2, AS1 does not benefit from the other peer-to-peer edges of AS2: a peer will not transit traffic for a peer. Therefore, high peer-to-peer degree does not make a node more attractive as a peer-to-peer neighbor. We intend to investigate the validity of this explanation in the future.

2) *Clustering Coefficient*: *Clustering coefficient* is a metric that has been used to characterize and compare generated and real topologies [25]. Intuitively, the clustering coefficient captures the extent to which a node’s one-hop neighborhood is tightly connected—it is the ratio of the number of edges that the neighbors of a node have among themselves over the total possible number of such edges. For a node v_i with $n_i > 1$ neighbors, the clustering coefficient of v_i is $\gamma_i = \frac{m}{m_{\max}}$, where $m_{\max} = \frac{n_i(n_i-1)}{2}$, and m is the number of edges between these neighbors. A clustering coefficient of exactly one means that the neighborhood is a clique. The average clustering coefficient of *OBD* is 0.25 and it increases to 0.31 in *ALL*.

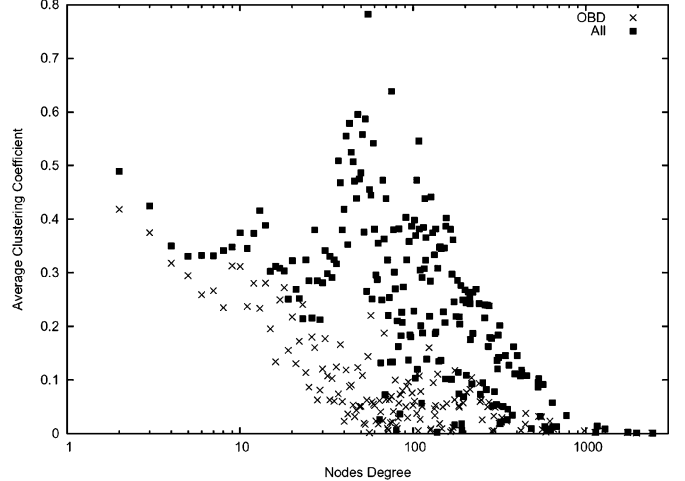


Fig. 4. The per-degree average clustering coefficient $\bar{\gamma}_d$ versus the degree for graphs *ALL* and *OBD*.

In addition, we find that the density increase is not homogeneous. The neighborhoods of “middle-class” nodes become more clustered. We use $\bar{\gamma}_d$ to denote the *average clustering coefficient* of all nodes with degree d . In Fig. 4, we plot $\bar{\gamma}_d$ versus the node degree d , for two graphs: *ALL* and *OBD*. The *ALL* graph has overall higher clustering coefficients as expected. We find that the clustering coefficient increase is larger for nodes with degrees in the 10 to 300 range. Note that this property characterizes the new edges, and could help us identify more missing edges in future studies.

3) *AS Path Length*: We study the effect of the new edges on the AS path lengths with policy-aware routing. The routing policy is a consequence of the business practices driven by contracts, agreements, and ultimately profit. As a first-order approximation of the real routing policy, we use the *No Valley Prefer Customer (NVPC)* routing, as described in [32], [33].

We have approximately 20 000 ASes present in the Internet topology and examine all possible pairs of ASes. For each AS pair, we compare the AS path lengths with *OBD* and with *ALL*. We group those AS paths with the same shorter path length, and show their path length changes in Fig. 5. We find that approximately 10 million paths change in length. While we note that this is a small fraction of the total number of paths, it is still a significant number in terms of its absolute value. In addition, *no change in the length does not mean that the path did not change*. For this reason, we study how many paths changed even if they did not change in length in Section IV-C.

One interesting observation here is that, by discovering the new edges, some of the new paths become in fact longer than before! This would never happen if the routing policy was based on the minimum hop criterion. Here, the length increase is due to the routing policy, which is based on the business relationship types. In other words, an AS will prefer a longer path through a peer than a shorter path through its provider.⁵ In Fig. 6, we show all possible changes that a new peer-to-peer edge (AS 2–AS 3) can cause to a path (from AS 4 to AS 5): (a) shorten the path;

⁵In practice, this is done by setting higher *local-pref* value to peer links than to provider AS links. When multiple BGP paths to a prefix are available, BGP will first choose the route with the highest *local-pref* value. [34]

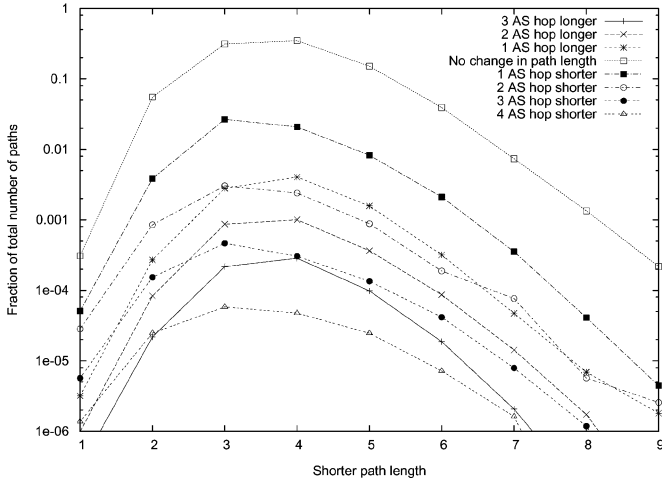


Fig. 5. The effect of the new links on the path length: fraction of the number of paths that change length versus the length of the shorter path, with each line representing a length change.

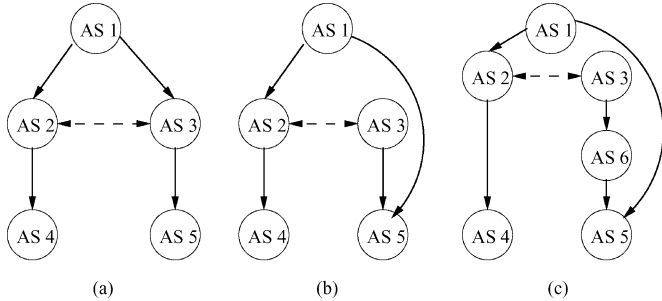


Fig. 6. The effect of adding a peer-to-peer link between AS 2 and AS 3 on the path from AS 4 to AS 5. The arrow points from the provider to the customer.

(b) change the path but maintain the same length; and (c) elongate the path. In more detail, Fig. 6(c) shows that initially the path from AS 4 to AS 5 is [AS4, AS2, AS1, AS5]. With a new peer-to-peer edge between AS 2 and AS 3, the path changes to a longer one [AS4, AS2, AS3, AS6, AS5]. The reason is that, in Fig. 6(c), AS 2 prefers to route through its peer AS 3 rather than through its provider AS 1 according to the No Valley Prefer Customer routing.

C. The Effect on ISP Revenue

We examine how much the new discovered AS links would change the models previous studies had arrived at about routing decisions and ISP income by using incomplete Internet topology.

Similar to studying AS path length, we assume NVPC routing in our model. For each AS, we count how many of its paths stop going through one of its providers once the new edges are added. We refer to these paths as *ex-provider paths*. The number of *ex-provider paths* is an indication, of the financial gains for that AS. Clearly, there are other considerations, such as prefix-based traffic engineering and performance issues, that our analysis cannot possibly capture. However, our results are a good first indication of the effect of the new peer-to-peer links.

The significant financial benefits of the new peer-to-peer edges. We plot the number of *ex-provider paths* for each node in

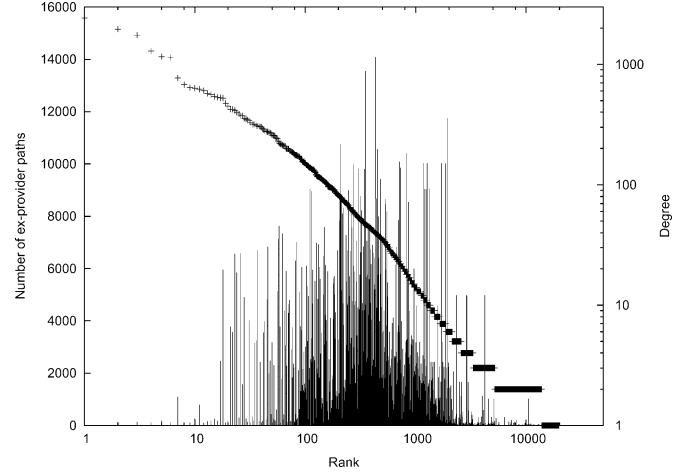


Fig. 7. The number of *ex-provider paths* (shown as impulses on the left y-axis) of each node in order decreasing node degree (shown as a semi diagonal line corresponding to the right y-axis). The x-axis shows the rank of the nodes in the order of descending degree.

Fig. 7. The x-axis represents the rank of the nodes on a log scale in order of decreasing degree; The y-axis at the left represents the number of *ex-provider paths*. In addition, we plot the node degrees (on the right y-axis) against their ranks as a semi diagonal line. Here we show an example of how to read the graph: nodes of rank 1000 (x-axis) correspond to nodes of degree approximately 10 (right y-axis) and have up to 12K *ex-provider paths* (left y-axis). We see that the difference between using an incomplete graph (*OBD*) and using a more complete graph (*ALL*) is dramatic: there are many ASes, for each of which, several thousands out of the total 20K paths (to all other ASes) stop going through a provider. For some ASes, more than 50% of their paths stop going through their providers (10K out of 20K possible paths per AS).

The rise of the “middle class” ASes. Another interesting observation is that the nodes which seem to benefit the most from these changes have degrees in the range from 10 to 300 (right y-axis). Top tier nodes (top 20 ranked) do not benefit almost at all. This is not surprising, since they do not have any providers anyway. Nodes with really low node degree do not benefit much either. One possible explanation is that these nodes do not have a lot of paths passing through them, as they don’t have many customer ASes.

D. Are We Missing a Lot More Peer Edges?

Currently, the *ALL* graph has approximately 20.9K peer-to-peer edges. However, we were very conservative in adding edges from *IRRnc*: we required that the edges are verified by *RETRO*. So, a natural question is, how many more edges could we verify from *IRRnc* if we had more *RETRO* servers. In other words, how many edges could we be missing? We attempt to provide an estimate by extrapolating the success of our method in finding new edges. Given the results above, we expect that the new edges would be of the peer-to-peer type.

Conservative estimate using *IRRdual*: We revisit the *IR-Rdual* graph and examine if we can include more edges than the ones we validate with *RETRO*. As shown in Table VII, there are 13 905 edges in the *peerIRRdual-BD*, and from these, only 4487 are “verifiable” candidates. Using *RETRO*, we verify 3529

or 78.6% of the verifiable edges. Here, we generalize this percentage: we assume that if we had more RETRO monitors, we could verify 78.6% of the 13 905 edges in *peerIRRDual-BD*. This leads to an estimated 7.4K (10.9K–3.5K) more peer-to-peer edges on top of the 20.9K peer-to-peer edges we currently have in ALL. Assuming that the estimate is correct, in our ALL graph, we are missing approximately 26% (7.4 out of 20.9+7.4) of the total number of peer-to-peer edges, or 35% of the total number of peer-to-peer edges that we have now.

Liberal estimate using IRRnc: In a similar way, we estimate how many edges we could verify from *peerIRRNc-BD*, which is a more “inclusive” set, that may contain more errors. Here, the total number of peer-to-peer edges is 39 894, the verifiable edges 8791, and the verified edges 5646. This gives rise to an estimate of $39\,894 \times 5646 / 8791 = 25.6\text{K}$ peer-to-peer edges out of which 5.6K are already in ALL. In other words, we have 20K new peer-to-peer edges on top of the current 20.9K peer-to-peer edges. Thus, in this more liberal guess, we may be missing 49% of the total peer-to-peer edges.

V. IDENTIFYING IXP PARTICIPANTS

In this section, we present a method for identifying the *participants* at Internet Exchange Points (IXPs). Our goal is to find all the participants at each IXP, and this is a nontrivial problem.⁶ We find that knowing the IXP participants is key for identifying many missing AS edges as explained in Section III.

Our approach consists of two complementary mechanisms: a technique to infer IXP participants using the IXP’s IP addresses, and an automated tool to parse and retrieve public archival information.

A. From the IP Addresses of IXPs

This part of our approach uses two techniques to infer IXP participants from IXP IP addresses: 1) path-based inference, where we perform a careful processing of collected traceroute data, and 2) name-based inference, where we analyze the name and the related information with regard to IXPs from the DNS and/or WHOIS databases.

In both inference methods, we start with the IP address blocks allocated to the IXPs, which we call *IXP IP addresses*. We obtain this information from the Packet Clearing House (PCH) [35]. In terms of traceroute data, we use a full cycle of Skitter traceroute data between May 1, 2005 and May 12, 2005, and our *RETRO_TRACE1* data in May 2005 as described in Section III-D.

1) *Path-Based Inference:* The high level overview of the method is deceptively simple. First, for each IXP IP address IP_{ixp} that we obtain from PCH, we search for the IP address that appears immediately after IP_{ixp} in each of the obtained traceroute paths. Second, if we find more than one such IP addresses for the particular IP_{ixp} , we select the one that appears most to be IP_{next} . We call the above procedure the *majority selection process*. Third, we find the AS AS_x that owns the IP address IP_{next} , and consider that AS_x to be a participant at the IXP.

⁶Efforts in improving IP-to-AS mapping try to identify IXP IPs, rather than the participant ASes. Their goal is different: they only need to find whether an observed IP address belongs to an IXP or not [19], [20].

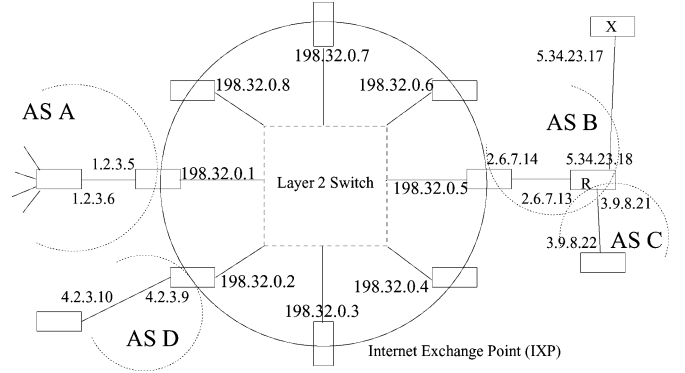


Fig. 8. Typical structure of an IXP.

Furthermore, we consider that IP_{ixp} is the IP interface via which AS_x accesses the IXP.

To illustrate this with an example, let us consider Fig. 8. A typical traceroute from AS A to router X yields the following sequence of IP addresses: [1.2.3.5, 198.32.0.5, 2.6.7.13, 5.34.23.17]. Since the address “2.6.7.13”, which belongs to AS B, appears immediately after IXP IP address “198.32.0.5”, we infer that AS B is a participant AS, and that 198.32.0.5 is the interface that is assigned to AS B. Note from Fig. 8 that, irrespective of the location of the traceroute source and its destination, if an IXP address (the address 198.32.0.5 in our example) appears in a traceroute, the IP address that appears immediately after (the address 2.6.7.13 in our example) is owned by the AS (in our example AS B) that uses the IXP address (e.g., 198.32.0.5) to access the IXP as long as two conditions hold. These conditions are: 1) each IXP interface address is assigned to a single AS, and 2) routers *always* respond to a traceroute probe with the address that corresponds to the incoming IP interface.⁷ While the first condition largely holds, the second condition does not always hold. There is a chance that a router could respond to a traceroute probe with an alternate (not the incoming) interface [19], [36]. In our example, router R could respond to a traceroute probe from AS A to router X with an alternate interface (e.g., 3.9.8.21), which makes the traceroute path appear as [1.2.3.5, 198.32.0.5, 3.9.8.21, 5.34.23.17]. Since 3.9.8.21 could be within the IP space of AS C, one could incorrectly infer that AS C is an IXP participant. We overcome this limitation with our *majority-selection process*; the basis is the assumption that *in the majority of the cases*, routers will respond to a traceroute probe with the incoming interface. This assumption has been shown to hold by numerous prior efforts [19], [36], [37].

The previously proposed method in [11] does not have the majority selection process. Furthermore the method does not associate the specific IXP IP interface addresses with their respective participating ASes. Our majority selection process eliminates measurement noise and thus, ensures a lower “false positive” rate. We map the discovered AS participants to their assigned IXP IP addresses, and using this, exclude the addresses in the name-based inference process that we describe below. This practice reduces the number of total IXP IP addresses that are

⁷The *incoming interface* of a traceroute probe is the IP interface via which the probe enters the router.

subject to the name-based inference procedures which are inherently less reliable, and thus reduces the possible errors overall.

2) *Named-Based IXP Participants Inference*: The basic name-based IXP participants inference method, which was proposed in [11], works in three main steps: 1) for every IP address in each IXP prefix space, we do a reverse DNS lookup, and we find the host name for that IXP IP address; 2) we take the domain name part (*company*.{com, net, org, etc.}) from the host name, and do a DNS lookup, which leads to a new IP address; and 3) we find the AS that owns this address, and this AS is considered a participant of that IXP. For example, IXP DE-CIX has the IP address 80.81.192.186. If we do a reverse DNS lookup, we get the host name

“GigabitEthernet3-2.core1.ftfl.level3.net”. A DNS lookup of the domain name “level3.net” yields an IP address of 209.245.19.41. An IP address to AS number conversion reveals that the IP address belongs to AS3356 (Level3). Therefore, AS3356 is considered a participant at DE-CIX.

Although this method has been used successfully by previous studies [11], it has two limitations: 1) sometimes it can return incorrect AS numbers for IXP participants,⁸ and 2) it does not always work: the DNS or the reverse DNS lookup may not return any answer.

We address the first limitation by excluding the IXP addresses that have been mapped on to AS participants by our path-based inference method. This greatly reduces the number of IXP addresses that are to be examined by the named-based inference method and therefore reduces the possible number of erroneous results.

We address the second limitation by proposing three new methods to improve the success rate of name-based inference:

1) *Examining host names containing AS numbers*. Sometimes, the DNS name of an IXP IP address contains the AS number of an IXP participant. For example, 195.66.224.71 is an IP address at the London Internet Exchange (LINX), which has a DNS name fe-3-4-cr2.sov.as9153.net. From that, we can infer that AS9153 is a participant at the LINX IXP.

2) *Examining common naming practices*. We can increase the success rate of DNS lookups by including common host names with the inferred domain names. For example, although *company.net* may fail to be resolved, the DNS lookup may succeed with *ns.company.net*. In fact, there are several common hostnames such as “ns”, “ns1”, “mail” and “www”. Hosts with these names *usually* belong to the same AS. For example, 195.66.226.104 is an IP address at IXP LINX at London, England. The host name of that IP address is “linx-gw4.vbc.net” and the DNS lookup for the domain name “vbc.net” is unsuccessful. However, the DNS lookup for “ns.vbc.net” returns the address 194.207.0.129, which belongs to AS8785 (Astra/Eu-X and VBCnet GB).

⁸Often the incorrectly reported participant AS number has a relationship with the correct one, e.g., they belong to the same company. For example, we use this method to examine the CERN IXP at Geneva, Switzerland. The method suggests erroneously that AS7018 (AT&T WorldNet Services) is a participant. On the contrary, AS2686 (AT&T Global Network Services) is one of the CERN IXP’s participants.

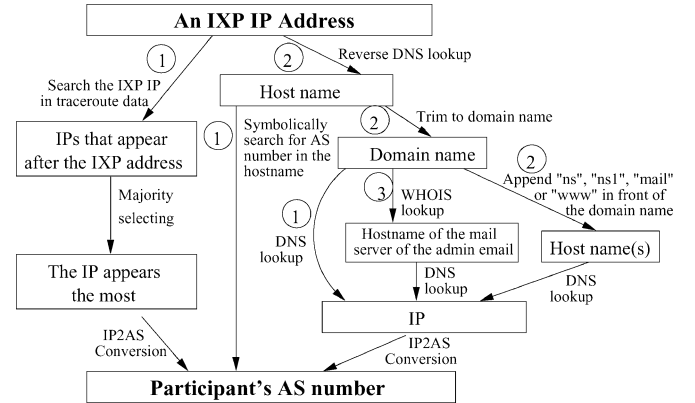


Fig. 9. The flow chart of our path-based method to infer IXP participants from IXP IP addresses. Starting from the top, the numbers in the circle indicate the priority (lowest number with highest priority) at a branching point.

3) *Using the administrating personnel information*. A WHOIS lookup for a domain name often has an administrative/technical contact person’s e-mail address. The mail server is often within the same AS that corresponds to the domain name. For example, for “decix-gw.f.de.bcc-ip.net”, all DNS lookups described previously, fail. However, if we look at the WHOIS lookup for domain “bcc-ip.net”, we will find the contact e-mail server is “bcc.de”, which has an IP address of 212.68.64.114, and it belongs to AS9066 (BCC GmbH).

3) *Putting the Two Techniques Together*: We integrate both the path-based and named-based techniques, into a tool for inferring IXP participants from IXP addresses. We start with the path-based technique, and for every IP address in the IP block of an IXP, we try to find it in a traceroute path. If this works, then we do not re-examine this IP address. Otherwise, we use the name-based inference and we utilize the three mechanisms that we proposed above. For completeness, we show the flow chart of the inference method in Fig. 9.

4) *Evaluating Our Inference Approach*: We use two complementary metrics: *Recall* \mathcal{R} and *Precision* \mathcal{P} , which are widely used in the data mining literature for similar tasks. They are defined as follows: $\mathcal{R} = \frac{\mathcal{N}_{\text{correct}}}{\mathcal{N}_{\text{actual}}}$ and $\mathcal{P} = \frac{\mathcal{N}_{\text{correct}}}{\mathcal{N}_{\text{inferred}}}$ where $\mathcal{N}_{\text{correct}}$ is the number of correctly inferred participants from among those inferred, $\mathcal{N}_{\text{actual}}$ is the actual number of participants, and $\mathcal{N}_{\text{inferred}}$ is the total number of inferred participants. Note that the Precision metric, \mathcal{P} , has not been used in previous studies although it is critical for detecting false positives. Otherwise, we favor overly aggressive inference methods that suggest a large number of correct and incorrect participants.

For the comparison and for lack of a better criterion, we select the six largest IXPs (in terms of number of participants) for which we know the participants through the EURO-IX site [38] or the IXPs’ own web sites as of May 12, 2005. In Table IX, for each IXP, we list its actual number of participants, the number of ASes that our algorithm inferred, and the number of ASes that our algorithm inferred *correctly*. We also show the Recall and Precision metrics.

It is easy to see that: a) our approach is very effective in determining most of the participants in these IXPs, and b) our ap-

TABLE IX
IXP PARTICIPANTS INFERRING COMPARISON (AS OF MAY 2005)

Name of IXP	Full Name	Actual participants	XDZC Approach [11]				Our Approach			
			correctly inferred	total inferred	\mathcal{R}	\mathcal{P}	correctly inferred	total inferred	\mathcal{R}	\mathcal{P}
MSK-IX	Moscow Internet Exchange	154	90	115	68%	90%	136	156	88%	87%
JPIX	Japan Internet Exchange	110	58	82	53%	71%	107	128	97%	84%
FREEIX	Free Internet Exchange (France)	101	38	39	38%	97%	64	65	63%	98%
AMS-IX	Amsterdam Internet Exchange	211	177	220	84%	80%	182	200	86%	91%
LINX	London Internet Exchange	175	164	242	94%	68%	168	193	96%	87%
DE-CIX	Deutscher Commercial Internet Exchange	144	111	124	77%	90%	137	142	95%	96%

proach identifies correctly more participants than XDZC and almost always with better Precision. For the case of MSK-IX, we only have slightly lower Precision (by 3%) but a significantly higher Recall (by 20%).

Note that for the evaluation we need to use IXPs for which we have the ground truth (published participants). The IXPs we use here are rather large, and typically follow naming conventions, which is important for some of our heuristics. For IXPs that do not follow naming conventions, often the smaller IXPs, our method may not outperform as much the previous method [11]. It will be interesting to further compare the two methods in the future and isolate the effect of each heuristic on the total performance.

B. From Web-Based Archive

We notice there are some limitations on inferring IXP participants by the IXP IP addresses alone. For example, some IXPs may have turned off the “time_exceeded” ICMP error responses, and therefore, IXP IP addresses may be invisible by traceroute or appear as “*”s in responses to traceroute probes.

To overcome these limitations, we include an additional source of information by retrieving IXP participant information from the web sites. We have developed a tool that automatically downloads and parses the web pages, and outputs the AS numbers of the participants periodically. We use the European Internet Exchanges Association [38] which maintains a database with 35 IXPs and their participants. We are also able to collect information from the web pages of 31 other IXPs. Naturally, as any manually-maintained data, these archives can also contain inaccuracies. However, we did not find any major inconsistencies with our measured data.

C. The Combined Results

We applied our methods to infer the participants at various IXPs on May 12, 2005. We first use our web-based archival inference. For the rest of the IXPs, we collect information with regard to their IP address blocks from Packet Clearing House [35], and infer their participants from their IXP IP addresses by using our inferring heuristics. We identify 2348 distinct participants at 110 IXPs. Some ASes actively participate in multiple IXPs. For example, AS 8220 (Colt Telecom) is inferred as a participant in 22 different IXPs in 15 different countries. In this study, we have used the combined results as our source of IXP data.

VI. CONCLUSION

In a nutshell, our work develops a systematic framework for the cross-validation and the synthesis of most available sources of topological information. We are able to find and *confirm* approximately 300% additional edges. Furthermore, we recognize that Internet Exchange Points (IXPs) hide significant topology information and most of those new discovered peer-to-peer AS links are incident at IXPs. The reason for such a phenomenon is probably because, most missing peer-to-peer links are likely to be at the middle or lower level of the Internet hierarchy, and peering at some IXP is a cost-efficient way for the ASes to setup peering relationships with other ASes. We show that by adding these new AS links, some research results based on previous incomplete topology, such as routing decision and ISP profit/cost, change dramatically. Our study suggests that business-oriented studies of the Internet should make a point of taking into consideration as many peer-to-peer edges as possible.

So, how many AS links are still missing from our new snapshot of the Internet topology? Our findings suggest that if we know the peering matrix of all the IXPs, we might be able to discover most of the missing peer-to-peer AS links. Unfortunately, very few IXPs publish their peering matrices. Furthermore, the published peering matrices are not necessarily accurate, complete or up-to-date. In our conservative estimates, there might be still 35% hiding peer-to-peer edges, in addition to what we already have in current Internet AS graph.

Our future plans have two distinct directions. First, we want to continue the effort towards a more complete Internet topology instance. Using the framework we developed here, we are in a good position to quickly and accurately incorporate new information, such as new BGP routing tables, or new traceroute servers. Second, given our more complete AS topology, we are in a better position to understand the structure of the Internet and the socioeconomic and operational factors that guide its growth. This in turn could help us interpret and anticipate the Internet evolution and, indirectly, give us guidelines for designing better networks in the future.

REFERENCES

- [1] S. Floyd and V. Paxson, “Difficulties in simulating the Internet,” *IEEE/ACM Trans. Networking*, vol. 9, no. 4, pp. 392–403, Aug. 2001.
- [2] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, “Towards capturing representative AS-level Internet topologies,” *Computer Networks*, vol. 44, no. 6, pp. 737–755, 2004.
- [3] B. Zhang, R. A. Liu, D. Massey, and L. Zhang, “Collecting the Internet as-level topology,” *ACM SIGCOMM Comput. Commun. Rev. (CCR)*, vol. 35, no. 1, pp. 53–61, Jan. 2005.
- [4] Y. Shavitt and E. Shir, “DIMES: let the Internet measure itself,” *ACM SIGCOMM Comput. Commun. Rev. (CCR)*, vol. 35, no. 5, pp. 71–74, Oct. 2005.

- [5] X. Dimitropoulos, D. Krioukov, and G. Riley, "Revisiting Internet AS-level topology discovery," in *Proc. Passive and Active Measurement (PAM) Workshop*, Boston, MA, 2005.
- [6] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. Claffy, and A. Vahdat, "The Internet AS-level topology: Three data sources and one definitive metric," *ACM SIGCOMM Comput. Commun. Rev. (CCR)*, vol. 36, no. 1, pp. 17–26, Jan. 2006.
- [7] H. Chang, S. Jamin, and W. Willinger, "To peer or not to peer: Modeling the evolution of the Internet's AS-level topology," in *Proc. IEEE INFOCOM*, 2006, 12 pp.
- [8] L. Colitti, G. DiBattista, M. Patrignani, M. Pizzonia, and M. Rimonini, "Investigating prefix propagation through active BGP probing," *Microprocessors Microsyst.*, vol. 31, no. 7, pp. 460–474, 2007.
- [9] R. Cohen and D. Raz, "The Internet dark matter—on the missing links in the AS connectivity map," in *Proc. IEEE INFOCOM*, 2006, 12 pp.
- [10] Internet Routing Registry. [Online]. Available: <http://www.irr.net>
- [11] K. Xu, Z. Duan, Z. Zhang, and J. Chandrashekar, "On properties of Internet exchange points and their impact on AS topology and relationship," in *Networking*, 2004, pp. 284–295.
- [12] G. Siganos and M. Faloutsos, "Analyzing BGP policies: Methodology and tool," in *Proc. IEEE INFOCOM*, 2004, vol. 3, pp. 1640–1651.
- [13] Y. He, G. Siganos, M. Faloutsos, and S. Krishnamurthy, "A systematic framework for unearthing the missing links: Measurements and impact," in *USENIX NSDI*, Cambridge, MA, Apr. 2007.
- [14] Oregon Routeview Project. [Online]. Available: <http://www.routeviews.org>
- [15] Ripe Route Information Service. [Online]. Available: <http://www.ripe.net/ris>
- [16] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie, "Sampling biases in IP topology measurements," in *Proc. IEEE INFOCOM*, 2003, vol. 1, pp. 332–341.
- [17] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore, "On the bias of traceroute sampling, or power-law degree distributions in regular graphs," in *Proc. STOC'05*, Baltimore, MD, May 2005, pp. 694–703.
- [18] R. V. Oliveira, B. Zhang, and L. Zhang, "Observing the evolution of Internet AS topology," in *Proc. ACM SIGCOMM*, 2007, pp. 313–324.
- [19] Z. Mao, J. Rexford, J. Wang, and R. Katz, "Towards an accurate AS-level traceroute tool," in *Proc. ACM SIGCOMM*, 2003, pp. 365–378.
- [20] Z. M. Mao, D. Johnson, J. Rexford, J. Wang, and R. Katz, "Scalable and accurate identification of AS-level forwarding paths," in *Proc. IEEE INFOCOM*, 2004, vol. 3, pp. 1605–1615.
- [21] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with Paris traceroute," in *Proc. ACM IMC'06*, Rio de Janeiro, Brazil, Oct. 2006, pp. 153–158.
- [22] E. Shir, Dec. 2005, Personal communication via e-mail.
- [23] B. Donnet, P. Raoult, T. Friedman, and M. Crovella, "Efficient algorithms for large-scale topology discovery," in *Proc. ACM SIGMETRICS*, Jun. 2005, pp. 327–338.
- [24] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with Rocketfuel," *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, Feb. 2004.
- [25] S. Jaiswal, A. Rosenberg, and D. Towsley, "Comparing the structure of power law graphs and the Internet AS graph," in *Proc. 12th IEEE Int. Conf. Network Protocols (ICNP'04)*, 2004, pp. 294–303.
- [26] S. Jin and A. Bestavros, "An empirical study of inherent routing bias in variable-degree networks," Boston Univ., Boston, MA, Tech. Rep., 2003.
- [27] J. Xia and L. Gao, "On the evaluation of AS relationship inferences," in *Proc. IEEE Globecom*, 2004, vol. 3, pp. 1373–1377.
- [28] Traceroute. [Online]. Available: <http://www.traceroute.org>
- [29] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," in *Proc. ACM SIGCOMM*, 1999, pp. 251–262.
- [30] A. Medina, I. Matta, and J. Byers, "On the origin of powerlaws in Internet topologies," *ACM SIGCOMM Comput. Commun. Rev. (CCR)*, vol. 30, no. 2, pp. 18–34, Apr. 2000.
- [31] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "The origin of power laws in Internet topologies revisited," in *Proc. IEEE INFOCOM*, 2002, pp. 608–617.
- [32] L. Gao and F. Wang, "The extent of AS path inflation by routing policies," in *Proc. IEEE Globecom*, 2002, vol. 3, pp. 2180–2184.
- [33] N. T. Spring, R. Mahajan, and T. E. Anderson, "The causes of path inflation," in *Proc. ACM SIGCOMM*, 2003, p. 113–124.
- [34] M. Caesar and J. Rexford, "BGP routing policies in ISP networks," *IEEE Network*, vol. 19, no. 6, pp. 5–11, Nov./Dec. 2005.
- [35] Packete Cleaning House. [Online]. Available: <http://www.pch.net>
- [36] L. Amimi, A. Shaikh, and H. Schulzrinne, "Issues with inferring Internet topological attributes," in *Proc. SPIE ITCOM*, 2002, vol. 4685, pp. 80–90.
- [37] Y. Hyun, A. Broido, and K. Claffy, Traceroute and BGP AS path incongruities. [Online]. Available: www.caida.org/outreach/papers/2003/ASP/
- [38] European Internet Exchange Assoc. [Online]. Available: <http://www.euro-ix.net>



Yihua He (S'04–M'07) received the Bachelor's degree from the South China University of Technology, China, the Master's degree from Kent State University, Kent, OH, and the Ph.D. degree at the University of California, Riverside, in 2007, all in computer science.

His research interests are in the field of Internet topology, routing protocols, measurements and evaluation. He is currently a technical member of Yahoo! Inc.



Georgos Siganos (S'01–M'04) received the Bachelor's degree from the Technical University of Crete and the Ph.D. degree from the University of California, Riverside.

He currently is with Telefonica Research, Barcelona, Spain. His research interests include, peer-to-peer systems, Internet Routing protocols, Internet Routing Registries and Internet measurements.



Michalis Faloutsos (M'99) received the Bachelor's degree from the National Technical University of Athens, Athens, Greece, and the M.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada.

He is a faculty member with the Computer Science Department at the University of California, Riverside. His interests include Internet protocols and measurements, network security, and routing in ad hoc networks. With his two brothers, he co-authored the paper "On powerlaws of the Internet topology" (SIGCOMM'99), which is in the top 20 most cited papers of 1999.

Dr. Faloutsos' work has been supported by several NSF and DARPA grants, including the prestigious NSF CAREER award. He is actively involved in the community as a reviewer and a TPC member in many conferences and journals. Recently, he has been authoring the popular column "You must be joking..." in the *ACM SIGCOMM Computer Communication Review*.



Srikanth V. Krishnamurthy (S'94–M'99–SM'07) received the Ph.D. degree from the University of California at San Diego in 1997.

From 1998 to 2000, he was a Research Staff Scientist at the HRL Laboratories, LLC, Malibu, CA. Currently, he is an Associate Professor of computer science at the University of California, Riverside. His research interests span wireless networks, sensor networks, Internet protocols and measurements, and network security.

Dr. Krishnamurthy has been a PI or a project lead on projects from various DARPA programs. He is the recipient of the NSF CAREER Award from ANI in 2003. He is the Editor-in-Chief of *ACM Mobile Computing and Communications Review* (MC2R).