The Missing AS Links and Their Impact On the Internet Topology Model

Yihua He
yhe@cs.ucr.edu
Georgos Siganos
siganos@cs.ucr.edu
Department of Computer Science and Engineering
University of California, Riverside

I. Introduction

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 is critical 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. We develop several methods and identify a significant number of missing links, particularly of the peer-to-peer type. In a nutshell, our study brings to light three major 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 appear in IRR, and (3) most of the missing AS edges are incident at IXPs.

In more detail, we make the following main contributions. First, we examine the Internet Exchange Points (IXPs) which have not been sufficiently explored in topology-oriented studies. We develop an improved method to identify IXP participants, and the edges among them, and find a significant number of peer-to-peer AS edges that do not appear anywhere else. Second, we conduct a cross-comparison and synthesis of most available sources of information: BGP routing tables, Internet Routing Registries, and traceroute data. In addition, we develop a tool to validate the existence of potential edges. Finally, we study the impact of the new edges.

II. On Finding Missing AS Links

In this section, we analyze and compare the information provided by different sources on the AS level topology.

We start with the Oregon routeviews BGP table dump (**OBD**), which is the most widely used data archive. We are interested in finding the edges that are missing from this dataset.

A. Use extra BGP routing tables:

We collect multiple BGP routing table dumps from various locations in the world, and compare them with **OBD**. On May 12, 2005, in addition to 6 BGP routing table dumps from the Oregon route collectors, we collect 28 more BGP table dumps from the RIPE/RIS route collectors and public route servers. We use the term **BD** to refer to the combined data from all 34 BGP table dumps. More than 99% edges in **BD** can be classified as either *provider-customer* edges or *peer-to-peer* edges by PTE [1].

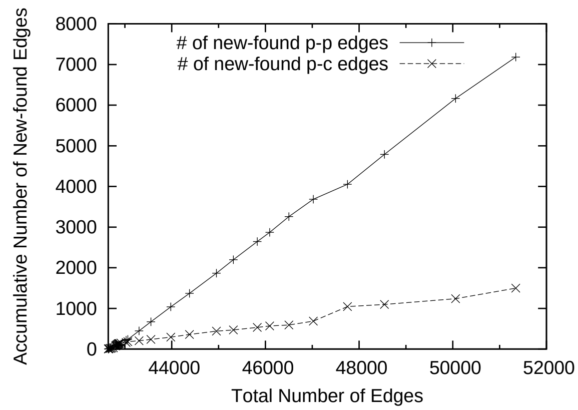

Fig. 1. Most of the new edges in BD but not in OBD are peer-to-peer edges.

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 Fig. 1, we find most of the new-found edges (the edges in *BD* but not in *OBD*) are of the peer-to-peer type.

B. Use selected IRR data to find potential edges:

We extract AS links from IRR on May 12, 2005 and classify their business relationships using Nemecis[2] as per the exporting policies of registered ISPs.

We make the following two observations. **a. IRR is a good source of hints for missing edges.** We perform the following thought experiment. *If we only knew of 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 Nemesis. 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, (in which the business relationships are inferred by PTE [1]) i.e., most new found AS edges are of the peer-to-peer type.

b. 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*. We find that 71.1% of the edges in IRR are still 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 are potentially many more IRR links that do exist but are yet to be verified.

C. Infer IXP participants and edges:

We start with a superset of the real IXP edges that contains all possible IXP edges: we consider the participants[1] of each IXP form a clique. We denote by **IXPall** the set of all edges that make up all of these cliques.

We consider two sets of data: (1)*peerIRR-BD*, the peer-to-peer IRR edges that not existing in *BD*; and (2)*peerIRRdual-BD*, similar to *peerIRR-BD*, but those peer-to-peer IRR edges are confirmed by exporting policies from both sides of the AS edges. With our first comparison, we find that approximately 60% of the edges in *peerIRR-BD* are in *IXPall* and hence, are potentially IXP edges. Furthermore, 83% of the *peerIRRdual-BD* edges are in *IXPall* thus, if they exist, they could be IXP edges. Interestingly, the *peerIRRdual-BD* is a more "reliable" subset of *peerIRR-BD*, and it exhibits a higher percentage of probable 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. Validate using RETRO:

We develop an essential tool, RETRO, for detecting and verifying AS edges by employing public traceroute servers. Currently, we have a total of 404 reverse traceroute servers which contain more than 1200 distinct and working vantage points. These vantages points cover 348 different ASes and 55 different countries.

To verify the existence of the edges in *peerIRR-BD*, we would like to witness these edges traceroute paths. For each edges in *peerIRR-BD*, we try to traceroute from one side of the edge to the other. We call **candidate** a potential AS edge for which we have a RETRO monitor suitable placed at one side of the potential AS edge and the other side is reachable by RETRO (not necessarily via one AS-hop). We have 8791 such "candidate" paths. By doing traceroutes appropriately, we get traceroute paths. Each of these paths includes the two ASes of interest, e.g. the ASes we think are connected with an edge. In these paths, we search for two patterns for each AS A and AS B in *peerIRR-BD*: (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 means that the AS edge between AS A and AS B exists as (a) a direct edge or, (b) as an IXP edge respectively. The results are summarized in Table I.

[1]We use an improved IXP participants inference algorithm based on [3].

TABLE I

RETRO VERIFIES PEER-PEER LINKS IN IRR MISSING FROM BD

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

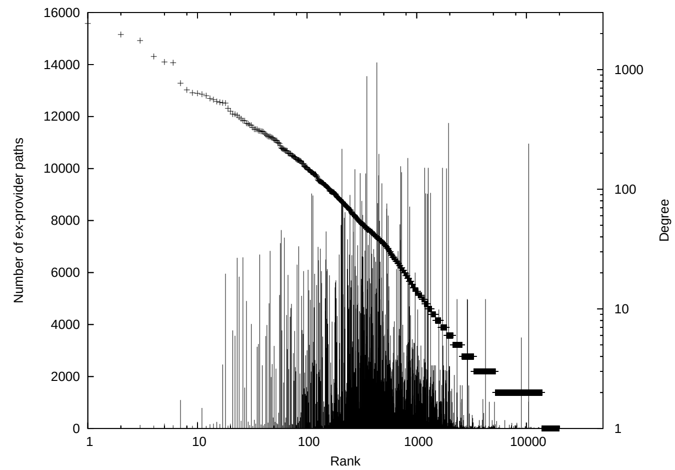

Fig. 2. The number of ex-provider paths (shown as impulses corresponding to the left y-axis) of each node in order decreasing node degree (shown as a semi diagonal line corresponding to the right y-axis). The x-axis show the ranking of the nodes(highest degree node ranks first).

We are able to confirm that a total of 5646 edges indeed exist, out of the 8791 AS edges that RETRO is able to check. Furthermore, from among the 5646 verified edges, 5317 or 94.2% of them are confirmed as IXP edges. Note that this method can only confirm the existence, but not prove non-existence of an edge. It could very well be that that the traceroute does not pass through the right path.

III. THE IMPACT OF THE NEW EDGES

We follow a commercial characterization: an AS pays for sending traffic through its provider. How much do the new edges affect these decisions? We count for each AS 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*. We plot the number of the *ex-provider* paths for each node in Fig. 2. We see that the effect on the routing on individual ASes 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. Another interesting observation is that the nodes which seem to benefit the most from these changes have degrees in the range from 10 to 100 (right y-axis).

REFERENCES

[1] Jianhong Xia and Lixin Gao. On the evaluation of as relationship inferences. In *IEEE Globecom*, November 2004.

[2] Georgos Siganos and Michalis Faloutsos. Analyzing bgp policies:methodology and tool. In *Infocom*, 2004.

[3] Kuai Xu, Zhenhai Duan, Zhi-Li Zhang, and Jaideep Chandrashekar. On properties of internet exchange points and their impact on as tolology and relationship. In *Networking*, 2004.