DISCOVERING AND VALIDATING MISSING LINKS IN THE INTERNET TOPOLOGY

S. Ravichandran
Associate professor,IT
Sri Krishna Engineering College
Chennai, Tamil Nadu, India

A.Vijayaraj
Associate Professor, IT Department,
Saveetha Engineering College
Chennai-602 105, Tamil Nadu, India
satturvijay@yahoo.com

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 the 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 the work is to develop methodologies and tools to identify and validate such missing links between ASes. In this work, to 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 to find exist as peer-to-peer links at the Internet Exchange Points (IXPs). First, in more detail, to provide a large-scale comprehensive synthesis of the available sources of information. To cross-validate and compare BGP routing tables, Internet Routing Registries, and traceroute data, while to extract significant new information from the less-studied Internet Exchange Points (IXPs). To 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, to identify properties of the new edges and quantify their effects on important topological properties. Given the new peer-to-peer edges, to 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. IXPs(Internet Exchange Points) have not received attention in terms of Internet topology discovery, although they play a major role in the Internet connectivity.

Keywords: missing links, Internet Exchange Points, peer-to-peer, extract, topology, traceroute data.

I. INTRODUCTION

An accurate and complete model of the Internet topology is critical for future protocol design, performance evaluation, simulation and analysis. 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. 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 to 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, to design and implement a systematic framework for discovering missing links in the 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, to apply the framework and conduct an in-depth study of the importance of these new links, and improve an understanding of the Internet topology at the AS level. In more detail, the 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,trace route, and IRR. Second, the framework extracts significant new topological information from Internet Exchange Points (IXPs); such information is typically not used in topological studies. While prior work has proposed methods to identify participating ASes at IXPs, to study greatly extends their work and overcomes certain limitations.

II. PROBLEM STATEMENT

To 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 entry contains an AS path, which corresponds to a set of AS edges. Several sites collect tables from multiple BGP routers, such as Route view and RIPE/RIS. 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. 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 discusses in depth this limitation.

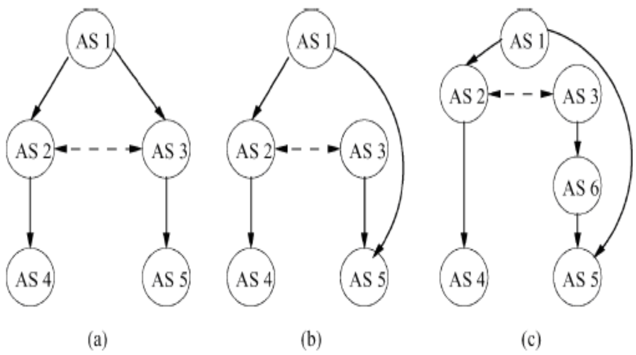


Fig .The effect of adding a peer-peer link between AS2 and AS3 on the path from AS4 to AS5. The arrow points from the provider to the customer

III. RELATED WORK

3.1 Analyzing BGP Policies

The overarching goal of this work is to model and improve the robustness of the Internet at the BGP level. The Border Gateway Protocol is the protocol that dictates routing between Autonomous Systems (AS), and implements their business policies. The importance of BGP has become clear in the network community over the last five years, and several efforts have improved understanding of BGP. However, to still have a long way to go: studies show that BGP operates in a far from robust state and many of its behaviors are not well understood. The need for a robust Internet has created efforts like the Internet Routing Registries (IRR), a distributed database, where ASes store their policies. However, IRR has not reached its potential nor fulfilled the initial vision. The work attempts to take the IRR to the next level. To provide a systematic approach and a tool, Nemecis, to extract and infer useful information from IRR, with the ultimate goal to use this information to model, manage and protect Internet routing. There exist a number of tools to measure actual BGP routing, like ping, traceroute, looking glass, BGP table dumps. Internet is structured into a number of routing domains that have independent administrations, called Autonomous Systems (AS). Each autonomous system is identified by a number, asn, which is assigned to it by an Internet registry. An Autonomous System uses an intra-domain routing protocol, like OSPF or IS-IS, inside its domain, and an inter domain protocol to exchange routing information with other Autonomous Systems. The defacto standard for inter-domain routing is BGP-4 The primary difference between the intra-domain and the inter-domain protocol is that the first one is optimized for performance, solely based on operational requirements, while the second is used to enforce the policy of the Autonomous System, which corresponds to the business relations with its neighboring ASes. An Autonomous System given its policy, will advertise to its neighbors a list of IP Prefixes, or routes that are reachable through it.

3.2 Scalable and Accurate Identification of AS-Level Forwarding Paths

Traceroute is widely used to detect routing problems, characterize end-to-end paths, and discover the Internet topology.Traceroute sends a series of TTL-limited probes toward a target destination, and reports the interfaces on the forwarding path and the round-trip time for each hop. In Figure 1, the _rst column shows the output of the traceroute to CNN's web site. This is invaluable to network operators and researchers. For example, network operators use traceroute to identify forwarding loops, blackholes, routing changes, unexpected paths through the Internet, and the end-to-end latency. Upon detecting a routing or performance anomaly, operators need to identify the Autonomous System (AS) responsible for the problem. The second and third columns of the Figure 1 denote the AS information of the routers along the forwarding path. Inaccurate information about the ASes along the path leads to delays in identifying and correcting the problem. In addition, research studies based on AS paths or graphs derived from traceroute depend on having an effective way to map the traceroute data to an AS-level forwarding path.

3.2.1 IMPROVING IP-TO-AS MAPPINGS

The dynamic programming algorithm, by producing an optimal matching for each pair, can also be used to improve the mapping by helping to identify places where the mapping is accurate. For example, the mapping from the BGP tables assigns pre_x 154.54.10.0/24 to AS2149, but in the 7972 pairs in which an IP address appears for which this is the longest pre_x , to optimal matchings never match it to AS2149, but match it to AS174 91.8% of the time (even though there is a penalty involved). This suggests that the actual mapping should have been to the latter AS. Indeed, if to change the
mapping accordingly, the optimal matchings for the relevant pairs now match 154.54.10.0/24 with AS174 100% of the time! In this paper, to consider a simple scheme for improving mappings by systematically exploiting the information from optimal matchings. Other schemes are possible, but thecurrent results already show significant improvement over the previous approach and provide a starting point for other more sophisticated rules.

3.3 A Systematic Framework for Unearthing the Missing Links: Measurements and Impact

An accurate topology model would be important for simulating, analyzing, and designing the future protocols effectively. With an accurate Internet AS-level topology, first, to can design and analyze new inter domain routing protocols, such as HLP that take advantage of the properties of the Internet AS-level topology. Second, to can create more accurate models for simulation purposes. Third, to can analyze phenomena such as the spread of viruses more accurately. In addition, the current initiatives of rethinking and redesigning the Internet and its operation from scratch would also benefit from such a model. Developing an accurate representation of the Internet topology at the AS level remains as a challenge despite the recent flurry of studies. Currently, several sources of topological information exist:

(a) archives of BGP routing tables, (b) archives of BGP routing updates, (c) Internet Routing registries, and (d) archives of trace route 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; these views are often complementary. Furthermore, as far as now, IXPs (Internet Exchange Points) have not received attention in terms of Internet topology discovery, although they play a major role in the Internet connectivity. The contributions of this work are two.

(a) To design and implement a systematic framework for discovering missing links in current Internet topology snapshot, and address two limitations of previous studies—the synthesis of different data source and incorporating topological information from IXPs. (b) To apply the framework and conduct an in-depth study of the importance of these new links, and improve the understanding of the Internet topology at the AS level.

IV. PAPER DESCRIPTION

4.1 Framework for Finding Missing Links:

To 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, cannot just simply take the union of all the edges. A careful synthesis and cross-validation is required. At the same time, these are interested in identifying the properties of the missing AS links.

4.2 New Edges From a BGP Table Dump :

To collect multiple BGP routing table dumps from various locations in the world, and compare them with OBD. For each BGP routing table dump, to extract its "AS PATH" field and generate an AS topology graph. Then merge these 34 graphs into a single graph and delete duplicate AS edges if any.

Table

TOPOLOGICAL DATA SETS USED IN OUR STUDY

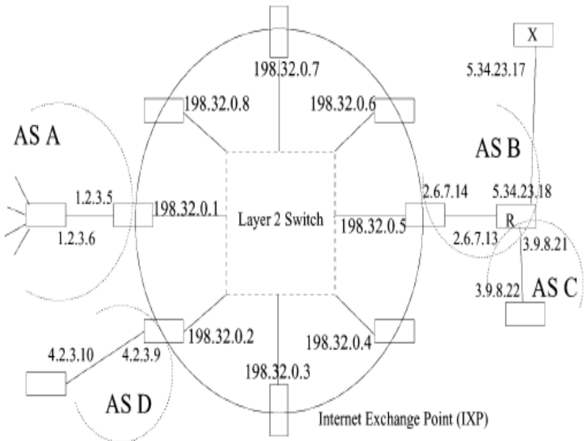
OBD	The Oregon routeviews BGP table dump
BD	OBD and other additional BGP table dumps
IRRnc	IRR edges processed by Nemecis with non-conflicting policy declarations
IRRdual	IRRnc edges correctly declared by both adjacent ASes
BD+IRR	BD and the edges of IRRdual confirmed by RETRO
IXPall	Union of cliques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

4.3 Exploring IRR

To carefully process the IRR information to identify potential new edges. Recall that do not add any edges until to verify them with RETRO. 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.

4.4 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 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) to need to find the IXP participants, and 2) to need to identify which edges exist between the participants.



.

Fig.Typical structure of IXP

4.5 Validating Links With RETRO

With the work so far, have identified sets of edges and obtained hints on where to look for new edges: 1) most missing 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 have noted before, the peer-to-peer edges learned through the IRRs and IXP all are not guaranteed to exist. Therefore, to focus on validating their existence to the extent possible. Note here that with the validation, to eliminate stale information that may still be present in the IRR and IXP data sources. Most newfound peer-to-peer links are incident at IXPs. To define a candidate to be a potential edge between two ASes, which satisfy the following two conditions: 1) to 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. To have 8791 such “candidates” for the potential AS edges in peerIRRnc-BD. By appropriately performing traceroutes on candidates, to get traceroute paths. In these paths, to search for two patterns for each candidate (,): a) and b) . If either of the two patterns appears, it is almost certain that the AS edge between and exists either as a) a direct edge or b) as an IXP edge, respectively. The results that to obtain at the end of the above process are summarized.

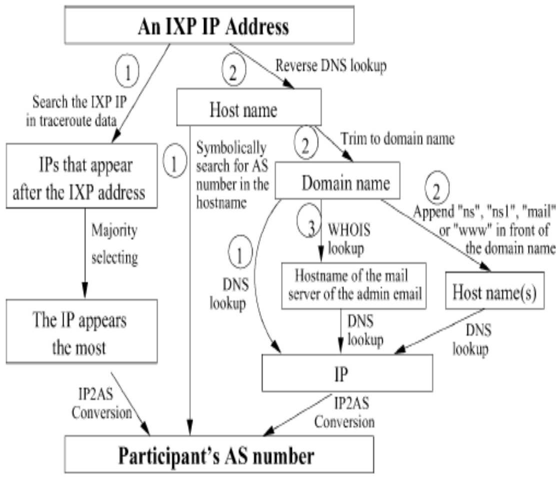


Fig : The flow chart of path-based method to infer IXP participants from IXP IP addresses. Starting from the top, the numbers in the circle indicate the priority (lowest number with highest priority) at a branching point.

V. FUTURE WORK

The future plans have two distinct directions. First, to want to continue the effort towards a more complete Internet topology instance. Using the framework is developed here, are in a good position to quickly and accurately incorporate new information, such as new BGP routing tables, or new trace route servers. Second, given the more complete AS topology, are in a better position to understand the structure of the Internet and the socioeconomic and operational factors that guide its growth.

VI. CONCLUSION

In this paper, I have proposed Discovering and Validating Missing Links in the Internet Topology. In a nutshell, the work develops a systematic framework for the cross-validation and the synthesis of most available sources of topological information. To able to find and confirm approximately 300% additional edges. Furthermore, to 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. To 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.

VII. REFERENCES

[1] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger,“Towards capturing representative AS-level Internet topologies,”

[2] Computer Networks, vol. 44, no. 6, pp. 737–755, 2004.

[3] 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. IEEEINFOCOM, 2006, 12 pp.

[4] L. Colitti, G. DiBattista, M. Patrignani, M. Pizzonia, and M. Rimondini

[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] S. Floyd and V. Paxson, "Difficulties in simulating the Internet,"IEEE/ACM Trans. Networking, vol. 9, no. 4, pp. 392–403, Aug. 2001.
[8] 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. IEEE/ACM TRANSACTIONS ON NETWORKING.
[9] B. Zhang, R. A. Liu, D. Massey, and L. Zhang, "Collecting the Internet
[10] as-level topology," ACM SIGCOMM Comput. Commun. Rev. (CCR), vol. 35, no. 1, pp. 53–61, Jan. 2005.

Author Profile:

A.Vijayaraj:-A.Vijayaraj is an Associate Professor in Department of Information Technology at Saveetha Engineering College. He received his Master of Computer Application in Bharathidhasan University, in 1997 and his Master of Engineering in Computer Science and Engineering from Sathyabama University at 2005. He has 12 years of teaching experience from various Engineering Colleges during tenure he was Awarded **Best Teacher Award** twice..He is a Member of, CSI and ISTE. He has Published 2 papers in International journal and 10 Papers in International and National Level conferences. His area of interest includes Operating Systems, Data Structures, Networks and Communication.

S. Ravichandran**:-** S. Ravichandran is an Associate Professor in Department of Information Technology at Sri Krishna Engineering College. He received his Master of Computer Application in Bharathidhasan University, in 1996 and his Master of Engineering in Computer Science and Engineering from Anna University at 2010. He has 15 years of teaching experience from various Engineering Colleges. He 6 Papers in International and National Level conferences. His area of interest includes cloud computing, artificial intelligence, compiler and Networks.