RESEARCH ARTICLE OPEN ACCESS

Generating Router Level Topology Using Dns And Ip Identifier

Rohini Ambure, Subodh Karve
Department of Computer Engineering Datta Meghe College of Engineering Airoli, Navi Mumbai
Department of Computer Engineering Datta Meghe College of Engineering Airoli, Navi Mumbai

ABSTRACT
Capturing an accurate view of the Internet topology is of great interest to the networking research community as it has many uses ranging from the design and evaluation of new protocols and services. But in real, topologies are not publicly available because ISPs generally regard their router-level topologies as confidential. So we define the steps for how to generate the router level topology within a minimum time with more accurate to the actual one. This paper describes the design of this steps and presents some preliminary analysis of the resulting router level topology.

Keywords- Router Level Topology; Internet Topology; Alias Resolution

I. INTRODUCTION

The Internet comprises thousands of independently-administered networks. Internet is a collection of thousands of smaller networks which are operated by Internet Service providers (ISP). Each ISP may have a different business model, and different high-level design. The independence of different networks means that network operators see and fix problems within their networks, but have little information about how other networks are run. Capturing an accurate view of the Internet topology is one of main interest to the networking research community as it has many uses ranging from the design and evaluation of new protocols. Internet topology maps are used for a vast number of applications, such as building models of the Internet, studying the robustness of the network, network management and improving routing protocols design. Hardware defects, or simply the reboot of a router can result in topology changes which in turn indirectly also result in routing changes. Router level topology maps are important to verify routing algorithms, in calculation and prediction of delay, node localization, traffic engineering, evaluating performance of P2P protocols, path restoration mechanisms, algorithms for building multicast trees and, in general, any study that would need a simulation over a realistic network scenario. The main contribution of this paper is to discover accurate router level topology by imposing the least possible overhead on the network within a least possible time. For this we are using different technique one, directed probing uses traceroute to identify the different route for topology. A second technique, path reduction reduces the number of path required to map by identifing redundant path.

The rest of this paper is organized as follows. In Section 2 and 3, we describe the existing tools for internet topology and our mapping techniques for router level topology. The implementation of our technique is described in Section 4. In Section 5 we are showing the generated topology after each technique. In Section 6 we are evaluating the generated topology.

II. EXISTING TECHNIQUE FOR INTERNET TOPOLOGY

The Mercator project [3] explores tools such as traceroute to group IP addresses in order to produce an Internet map. Mercator is also a map collection tool run from a single host. The aim of Mercator is to build a nearly complete map of the transit portion of the Internet from any location where Mercator is run, using hop-limited probing. Hop-limited probing differs from traceroute as it stops probing once a probe fails to elicit a response. This is appropriate for Mercator as it focuses on discovering router adjacencies. The technique used in Mercator is referred to as informed random address probing, in which a response from an IP address adds the address prefix to the Mercator list, and Mercator assumes that the neighboring prefixes are also addressable. Instead of a list of hosts, it uses informed random address probing to find destinations. Another assumption is sequential assignment of address space by the registries, such that, for example, 128.8/16 and 128.10/16 are the neighboring prefixes of 128.9/16. The CAIDA’s Skitter [4] project has been developed to combine *traceroute* and benchmark-based analysis. This tool uses *traceroute* to find the paths connecting two nodes and to collect performance information from them. Skitter uses BGP tables and a database of Web servers. The skitter data results in a spanning tree structure originating at the host and extending into the infrastructure toward the destination hosts. It then aggregate data into a centralized database for correlation and depiction as a top-down, macroscopic view of a cross-section of the Internet from at least a small set of sources. Skitter monitors probe the network from about 20 different locations worldwide. Donato Emma, Antonio Pescap´e, and Giorgio Ventre praposed hybrid IP based methodology for network topology discovery at router level[6]. Its implementation is in a tool called *HyNeTD* (Hybrid Network Topology Discovery) that effectively combines active and passive measurements to discover network topologies at router level. In [7] Spring present Internet mapping techniques that allow to measure directly router level ISP topologies without a significant loss in accuracy. The proposed techniques include the use of (i) BGP routing tables to focus measurements, (ii) IP routing properties , (iii) alias resolution techniques, and (iv) DNS queries. [12] presenting a problem related with identification of different interfaces, each interface with different IP address, that belong to the same router. It is vital in the process of discovering network topologies. They have analyzed the active probing and inference methods and have detected that one of the big problems is the filtering of replies to probing packets in routers. They have proposed modifications using different kinds of probing packets, trying in all of them to get more replies from routers for alias resolution. They have been improved the processing of the replies.

III. MAPPING TECHNIQUE

In this section, we present the steps for generating our topology. They are Directed Probing, Path Reduction, Alias Resolution and Router Identification.

3.1 Directed Probing

The very first step of topology generation is data collection. For our studies we will use the well known tool called traceroute by Jaconson whixh is used by several topology measurement. Traceroute running in alocal system provides the set of IP addresses of the routers in the path from source to destination system. It uses UDP probe packets with TTL field (Time To Live) starting from 1 and increasing one by one for each packet sent. TTL field will be decreased by one in each router in the path to destination, and when it gets 0 value the router will drop the packet and it will generate an ICMP packet with code “time to live exceeded in transit”. Repeating this procedure between several systems in certain networks we can get all the IP addresses of the routers in the topology.

3.2 Path Reduction

It is not possible that all the traceroute probes identified by directed probing will take unique paths inside the ISP. The number of measurements required can be reduced further by identifying probes that are likely to have identical paths inside the ISP. We list three different techniques here that have common properties of IP routing to cut down on redundant measurements. Ingress and egress redictions remove likely duplicates so that more valuable traces can be taken

3.2.1 Ingress Reduction

The path taken by a packet through a network is usually destination-specific. When traceroutes from two different vantage points to the same destination enter the ISP at the same point, the path through the ISP is likely to be the same. This is illustrated in Figure 2a. Since the traceroute from S2 to the destination would be redundant with the traceroute from S1, only one is needed. This redundancy can also be exploited to balance load between traceroute servers.

3.2.2 Egress Reduction

Similarly, traces from the single source to the two different destination, the path through the ISP is same. Such traces are redundant, so only one needs to be collected. This is shown in Fig 2b.

3.2.3 Next hop Reduction

The path through an ISP usually depends only on the next-hop AS, not on the specific destination prefix. This means that only one trace from ingress router to next-hop AS is likely to be valuable, as illustrated in Figure 2c.

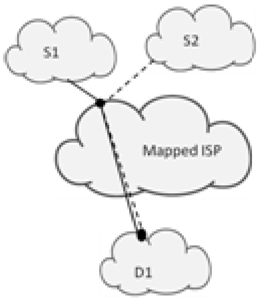

Fig 2a:Ingress Reduction

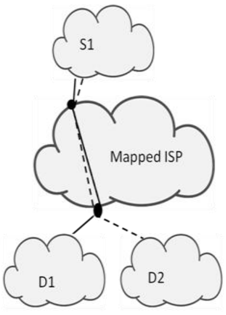

Fig 2b: Egress Reduction

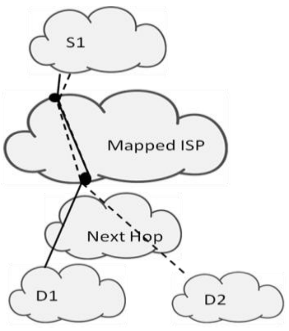

Fig 2c: Next hop Reduction

3.3 Alias Resolution

In order to provide topology information at router-level we need to identify different IP addresses belonging to the same router. Alias resolution will allow to reduce the set of expansion trees, coming from the traceroutes, to a network topology graph, with the process of reducing all the nodes that represent the same router to only one node with all its interfaces. These addresses are called alias and procedure for router identification is called as Alias resolution. In the Fig 3 traceroute lists the input interface 1 and 2 in left but the alias resolution technique gives the correct map. Interface 1 and 2 are alias. If the different addresses that represent the same router cannot be resolved, we get a different topology with more routers and links than the real one. Existing methods for alias resolution is categorized on 2 group: active probing methods and Inference methods[10]. Active probing techniques are based on sending specific probing packets to the routers and analyzing the replied packets. Inference methods try to deduce alias information by analyzing data from traceroute paths or by getting extra data from DNS.

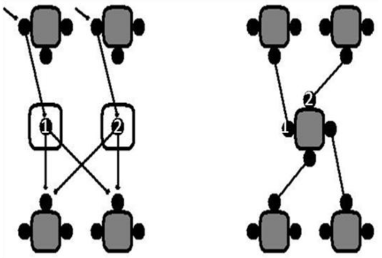

Fig 3: Alias Resolution

3.3.1 Based on source IP address

This method consists of sending UDP datagrams from the same host to all the IP addresses that could belong to the same router. The destination UDP port is chosen randomly and if there is no application listening on that port, the router sends an ICMP packet with type "destination port unreachable". These ICMP messages are generated from the interface with path to destination (the probing source host). So if all probed IP addresses belongs to the same router, all the ICMP messages will have the same source IP address and we will have a positive alias resolution. The method based on active probing is implemented by CAIDA [11] in the tool iffinder. However, nowadays it is usual to filter out those ICMP packets in the routers. so 92.55% of probing packets do not receive the ICMP notification because of filtering in the target router. Therefore, the effectiveness of this method is very limited.

3.3.2 Based on IP identifier counter

This is an alternative method of active probing which uses the same type of probes based on UDP packets to get ICMP notifications but in this case the method uses the IP identification field (IPID) in the IP header to check aliases. This IP identifier is originally used in the procedures of fragmentation and reassembly. This field has the same value for all fragments belonging to an original IP datagram before fragmentation, so it is used to reassemble the original packet in destination. Typical TCP/IP implementations of IP identifier use a counter which is incremented by one for each packet created in the host, independently of destination, protocol or service. Therefore, several IP packets received from the same host and near in time will have close values in the IP identifier field. The differences in the counter will be caused by other IP traffic generated in between by that host to other destination. This method was proposed in [7].This technique sends a probe packet to the two potential aliases. The port unreachable responses include the IP identifiers x and y. Ally sends a third packet to the address that responded first. If $x < y < z$, and $z - x$ is small, the addresses are likely aliases. As an optimization, if $|x-y| > 200$, the aliases are disqualified and the third packet is not sent.

3.3.3 Based on Graph analysis

The data collected by traceroutes can be used to construct a set of expansion trees (directed graphs) using the IP addresses as nodes and the pairs of IP addresses as edges. We will have an expansion tree for each source of the traceroute. We have to join information for several expansion trees with different sources and destinations in order to get a final graph with routers as nodes and links between routers as edges. Several heuristics for this process as

- Two addresses that immediately preceed a common successor are aliases[10].
- Different IP addresses that appear in the same traceroute trace cannot be aliases[10].
- Analytical Alias Resolver (AAR) [10] searches for potential path symmetry between two end points, locating point-to-point subnets (/30 or /31 networks) with IP addresses in each direction. If a match is observed, aliases can be found from the proper alignment of the path traces.

3.3.4 Based on DNS

In [7][12] the alias resolution scheme is based on drawing inferences from systematic naming conventions in DNS names. Usually ISPs follow a convention in naming interfaces of routers. Hierarchy embedded in DNS names and lexigraphically adjacency in name convention for router interfaces make it possible. Another application of DNS information is testing aliases as similarities in DNS names will identify alias pairs.

3.3.5 Based on IP Offset

To decide if two IP addresses have a certain probability for being alias, this method will use the offset between both IP addresses considered as two unsigned integer numbers of 32 bits. Basically the method will use the result of subtracting one IP address from the other (|IP1 −IP2|) to suggest the relation between them. The offset between two IP addresses will be called *IP offset*.

Active probing methods are characterized by introducing extra traffic in the network. They are intrusive so it is important to delimit the necessary injected traffic. Besides, this traffic can be confused with scanning or attacks, so many times they can have problems with filtering in routers.

3.4 Router Identification

In this step we describe how to determine which router in the traceroute output belong to the ISP, their geographical location in the topology.

3.4.1 Using DNS

We rely on the DNS to identify routers that belong to the ISP. The DNS names provide a more accurate characterization than the address space advertised by the AS for three reasons. First, routers of non-BGP speaking neighbors are often numbered from the AS's address space itself. In this case, the DNS names help to accurately locate the ISP network edge because the neighboring domain routers are not named in the ISPs domain (att.net, sprintlink.net, etc.). In some cases, the directly connected neighboring domain routers have a special naming convention that helps locate the network edge. For instance, small neighbors (customer organizations) of Sprint are named sl-neighborname.sprintlink.net, which is different from Sprint's internal router naming convention. Second, edge links between two networks could be numbered from either AS's address space. Again, DNS names help to identify the network edge correctly if they are assigned correctly. Finally, DNS names are effective in pruning out cable modems, DSL, and dialup modem pools belonging to the same organization as the ISP, and hence numbered from the same address space.

For example, sl-bb11-nyc-3-0.sprintlink.net is a Sprint backbone (bb11) router in New York City (nyc), and p4-0- 0-0.r01.miamfl01.us.bb.verio.net is a Verio backbone (bb) router in Miami, Florida (miamfl01).

IV. PROPOSED SYSTEM

In this section, we describe the details of the proposed system and implementation of the generated topology using the above techniques as shown in the Figure 4.

A database stores all the information which provides both persistentstorage of each measurements result and for interprocess communication between running process. The use of a database enables us to run SQL queries for simple question and allow us to integrate new analysis modules easily.

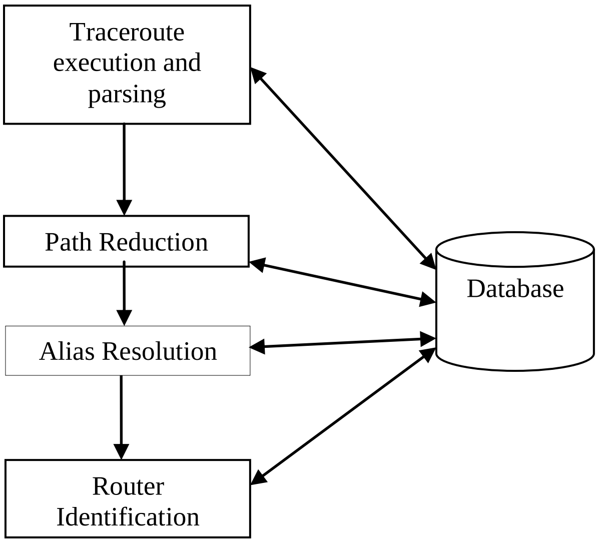

Fig 4: Architecture

We used existing traceroute command and extracts IP addresses that represent router interfaces and pair of IP addresses that represent links between them. We have started with 3000 destinations and used traceroute from single network node.

Path reduction take the data from the database identifies the duplicated IP addresses and links, reduces it with only one IP address and all the links are transferred with single IP address only.

Alias resolution based on Source IP address, IP Identifier Counter and IP offsets described in Section 3.3 are implemented, But the disadvantage of the above method is about how to guess two IP address pairs are aliases, for this we group the IP addresses according to their domain names which is available in the traceroute command's output.

Finally we are finding the geographic location of each router using longitude and latitude to show it on world map.

Algorithm for Directed Probing:

Create a text file URL.txt with list of URL for executing tracerout command.

Read the above file line by line until end of file

Retrieve URL and execute tracerout command

Store the tracerout output in the text file TraceOutput.txt

Read the above file line by line until end of file

For each Line

- Separate file line by two white space character and store it in array
- Insert the arrays last value which is IP address in one column of the row

Like this insert all the IP addresses in one row for one URL.

Reapeat the above steps for each URL from the URL.txt file.

Algorithm for Path Reduction:
For finding redundant edge and node first Create two recordset of the stored records during the directed probing.
For each IP address in first recordset

- Compare it to all the IP addresses in the second recordset.
- If the IP address is not found in second recordset after comparing with all IP address, Insert IP address with its x, y coordinate in the table.
- If IP address is found, stop comparing with all the records in second recordset.

Copy x, y coordinate of each IP address in new table from the table created during directed probing where IP addresses must be identical.

Algorithm for Aliases Resolution

1. **Based on IP Offset**

For two IP address IP1 and IP2 do if Ip1-IP2<=100
IP1 and IP2 are aliases otherwise not.

2. **Based on Source DNS**

For this method first we have to find the root domain from the domains which we have obtained from traceroute static-mum-59.185.210.198.mtnl.net.in, ge-6-0-0.0.cjr02.ldn001.flagtel.com these are the domains we are getting from active probing. From this first we separate root domain as mtnl.net.in, flagtel.com. After that we create one node for each domain and the IP address whose domain contains above substring creating a link from these node.

3. **Based on IP Identifier and DNS**

Alias resolution based on IP identifier is explained Chater 4, verifying all the possible pairs of IP addresses is difficult. So I have used domain names of the IP address obtained trought traceroute to verify the pair of IP address. But all the IP address don't have domain names in that case I have used IP offset method to check possibility that 2 IP addresses may be alias or not. After identifying pair of IP address, I am sending two ICMP packets to the IP addresses and obtained ICMP reply message and retrieved IP identifier field to check whether they are aliases or not.

Router Identification
For locating router on the world map we have used GeoLocateIP database from **IP2Location** [Appendix A] which is freely available on the internet. We used longnitude and latitude to place router on the world map.

V. GENERATED TOPOLOGY

In this section, we are presenting the samples of generated topology.

We have executed traceroute command for nearly 500 URL for collecting number of IP address. Figure 5 shows the topology from the data collected by probing. After that we have removed duplicate links and nodes from the graph obtained in Level 1, the resulting topology is shown in Figure 6. Figure 7 shows the topology by alias resolution based IP address offsets. Figure 8 shows the topology generated by alias resolution based on domain names.

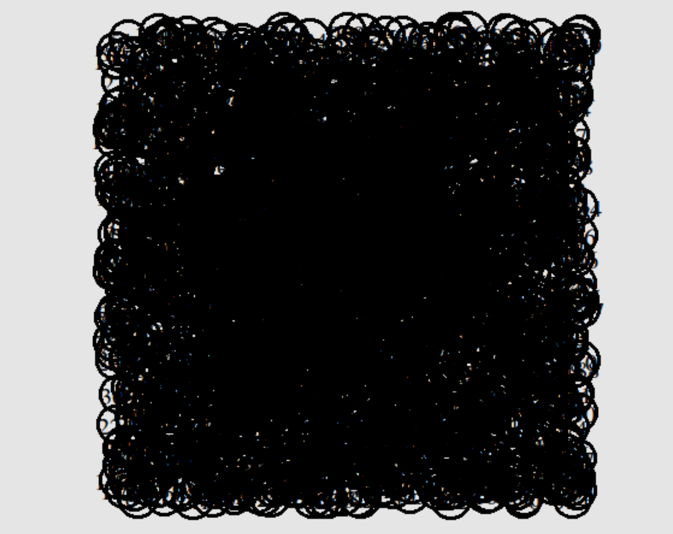

Fig 5: Level 0 Topology

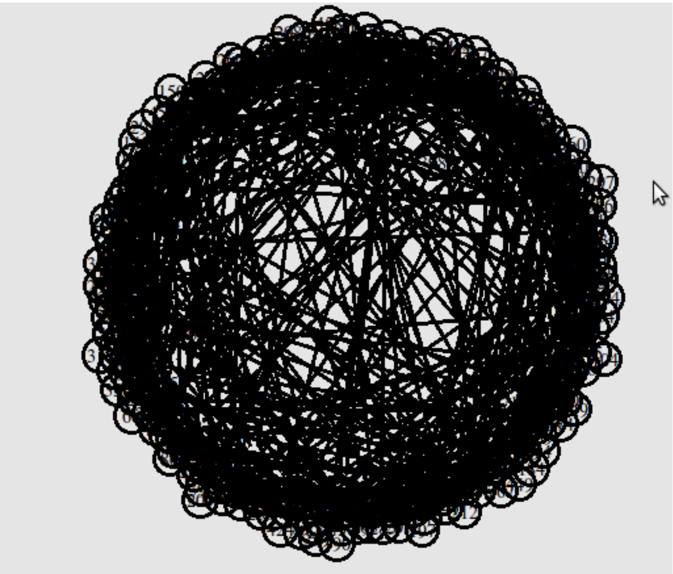

Fig 6: Level 1 Topology

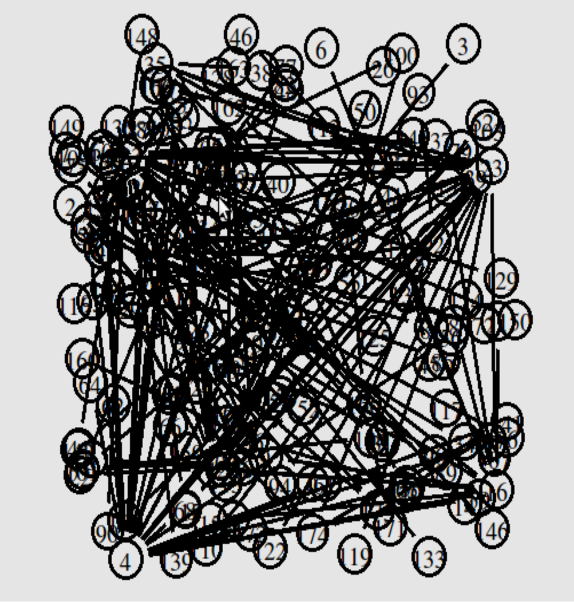

Fig 7: Level 3 Topology

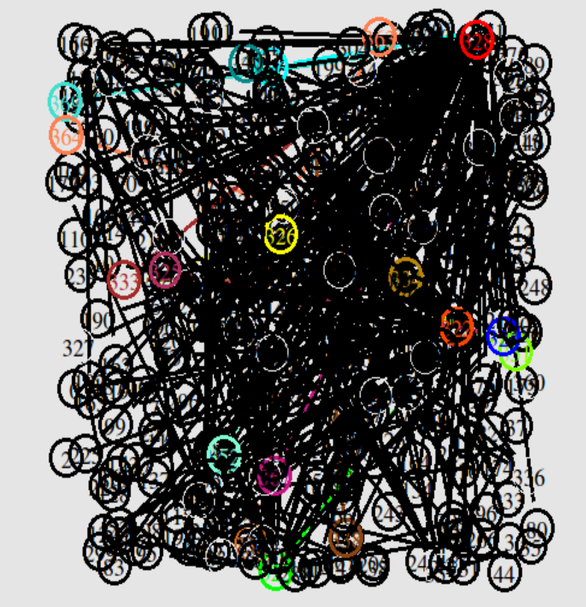

Fig 8 : Topology based on DNS

VI. EVALUATION

In this section we evaluate the effectiveness of our techniques based on two things: Completeness of the resulting maps and the efficiency with which they are constructed.

1. Completeness

We use the independent tests to estimate the accuracy and completeness of our maps. For this we ask the ISPs we mapped to help for validation.

Validating with ISPs

Two ISPs assisted us with validation of their maps. Below the list the question we asked and the answer we received from them.

Did we miss any router?
Did we miss any links between the router?
What fraction of routers did we miss?
Overall, do you rate our maps:

ii. Impact of Reduction

In this section, we evaluates the directed probing and path reductions described in Chapter 4. We evaluate these techniques for the extent of reduction in measurements that they bring.

Directed Probing

In this we consider, the number of traces that we should have kept. As we are using directed probing technique using traceroute command, every IP obtained from one traceroute is unique. So here we used all the IP addresses obtained from traceroute for thousands of IP addresses.

Path Reduction

In this section, we evaluate path reduction for its effectiveness in discarding unnecessary traces. Overall, ingress reduction kept only 17% of the traces chosen by directed probing.

VII. CONCLUSION

In recent years, a researcher in many different areas in networking has recognized the need for a random topology generator that produces realistic Internet topology. In order to produce realistic topologies, we need a better understanding of the Internet topology itself. This paper documents the several techniques to infer the Internet map, and reports on our generated result even though there are existing technique available such as Meracator, Skitter. For generating topology first we collected data by executing traceroute command for the number of URL. After that by applying path reduction with which we reduced our dataset to unique traces and various alias resolution methods we get the topology. Alias resolution based on IP identifier using Domain Names reduced more time require to guess that whether 2 IP interfaces are aliases or not. We identified the role and location of the router by using geographic information of IP address. The generated topology is almost accurate to the real topology.

Future Scope

We are planning to create a desktop network mapping service application like a Google map provided by Google. This network map will offer IP level topology, router level topology, Pop level topology and AS level topology across the world. At the root level user will get IP level topology like Bus, Ring, Star etc. At the top level user will get the network connectivity between the different countries. In short this will be the duplicate service of Google

Map only difference is Google map is based on street view, road map, rout plan and network map based on the connectivity between computer and networking devices.

References

[1] Topology, Hierarchy, and Correlations in Internet Graphs Romualdo Pastor-Satorras, Alexei V´azquez, and Alessandro Vespignani

[2] ftp://ftp.ee.lbl.gov/traceroute.tar.gz

[3] R. Govindan and H. Tangmunarunkit, "*Heuristics for Internet map discovery*", *Proc. of* IEEE Infocom '00, Mar. 2000.

[4] K. C. Claffy and D. McRobb. "*Measurement and Visualization of Internet Connectivity andPerformance*", http://www.caida.org/ TOOLS/measurement/skitter/ (As of July 2005).

[5] Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shenker, Walter Willinger "*Towards Capturing Representative AS-Level Internet Topologies*"

[6] Donato Emma, Antonio Pescap´e, and Giorgio Ventre, "*Discovering Topologies at Router Level* " University of Napoli "Federico II"

[7] N. Spring, R. Mahajan, and D. Wetherall, "*Measuring ISP Topologies with Rocketfuel*" *Proc. of* ACM/SIGCOMM '02, Aug. 2002.

[8] Yuval Shavitt and Noa Zilberman, "*Internet PoP Level Maps*"

[9] K. Yoshida, Y. Kikuchi, M. Yamamoto, Y. Fujii,K. Nagami, I. Nakagawa, and H. Esaki, "*Inferring PoP-level ISP topology through end-to-end delay measurement*," in Passive and Active Network Measurement. Springer, 2009.

[10] N. Spring, M. Dontcheva, M. Rodrig, and D. Wetherall, "*How to resolve IP aliases*," Tech. Report 04-05-04, Washington Univ. Computer Science, 2004.

[11] B. Huffaker, D. Plummer, D. Moore, and K. Claffy, "*Topology discovery by active probing*," in Proc. the Symposium on Applications and the Internet (SAINT), January 2002.

[12] S. Garc´ıa-Jim´enez, E. Maga˜na, D. Morat´o and M. Izal, "*Techniques for better alias resolution in Internet topology discovery*," Integrated Network Management, 2009. IM '09.

[13] David Moore, Colleen Shannon, Geoffrey M. Voelker, and Stefan Savage. Internet quarantine: Requirements for containing self propagating code. In Proceedings of the IEEE Joint Conference of the IEEE Computer and Communication Societies (INFOCOM), pages 1901-1910, San Francisco, CA, April 2003.

APPENDIX A

Founded in 2002, IP2Location is a Malaysian company offering IP geolocation software applications, i.e. tools that attempt to derive geographical data (country, region, city, latitude, longitude, ZIP code, time zone), and also connection speed, ISP and domain name, about an Internet user using their IP addresses. IP2Location.com is a subsidiary of Hexasoft Development Sdn. Bhd. ("HDSB"), a company based in Penang.