Issues with Inferring Internet Topological Attributes

Lisa Amini[a], Anees Shaikh[a], Henning Schulzrinne[b]
[a]IBM Research; [b]Columbia University

ABSTRACT

A number of recent studies are based on data collected from routing tables of inter-domain routers utilizing Border Gateway Protocol (BGP) and tools, such as `traceroute`, to probe end-to-end paths. The goal is to infer Internet topological properties. However, as more data is collected, it becomes obvious that data intended to represent the same properties, if gathered at different points within the network, can depict significantly different characteristics. While systematic data collection from a number of network vantage points can reduce certain ambiguities, thus far, no methods have been reported for fully resolving these issues. The goal of our study was to quantify the effect these anomalies have on key Internet structural attributes. We report on our analysis of over 290,000 measurements from globally distributed sites. We contrast results obtained from router-level measurements with those obtained from BGP routing tables, and offer insights as to why certain inferred properties differ. We demonstrate that the effect on some attributes, such as the average path length and the AS degree distribution can be minimized through careful data collection techniques. We also illustrate how using this same data to model other attributes, such as the actual forwarding path between a pair of nodes, or the level of AS path asymmetry, can produce substantially misleading results.

Keywords: Internet mapping, topology traceroute, Border Gateway Protocol, BGP

1. INTRODUCTION

The Internet's decentralized control and open nature have enabled it to evolve into an immense interconnection of millions of hosts, hundreds of thousands of address prefixes, and tens of thousands of separately administered routing domains. Measuring internal network parameters and mathematically modeling network structure is of significant importance. Such characterizations are used to isolate faults and pathologies within the Internet, improve existing protocols, validate proposals for new protocols, and predict the future evolution of the Internet.

Developing an accurate map of the Internet is challenging for a number of reasons. First, the structure of the Internet itself is not static – new nodes and edges are added daily. Secondly, as stated earlier, the number of nodes (hosts and routers) and edges (links) is enormous, and no single entity has complete knowledge of all Internet nodes and edges. Collaborative efforts, such as those by CAIDA [1], Oregon Route Views [2], and Looking Glass [3], have been established to acquire and share Internet traffic metrics. However, as more data is collected, it becomes obvious that data intended to represent the same properties, if collected with different tools or at different points within the network, can depict significantly different attributes. For example, a recent study of data collected from inter-domain routing tables led researchers to propose that the hierarchical structure of the Internet can be more compactly represented through power laws [4]. However, the accuracy of this model is now a topic of active debate [5].

In addition to the Internet's dynamic nature and decentralized administration, the most applicable view of the topology is not necessarily physical. More specifically, policy-based routing creates a logical overlay to the Internet's physical structure that determines how packets are actually forwarded. Because routing policies reflect business relations, network administrators frequently do not expose their routing policies.

Data currently being used for topology analysis is often obtained from the routing tables of the Internet's inter-domain routers, or by active, router-level probing techniques which generally rely on eliciting ICMP messages from remote routers or hosts [7]. A number of recent studies [5, 13, 14] have pointed out issues with utilizing this data to infer the Internet's structure. For example, since routers track reachability only from their location in the network, utilizing the routing tables of a single router, or a limited set of routers, will not provide a complete representation of Internet topology [5, 14]. Router-level probing provides additional structural details, but if collected from a limited set of network vantage points, will also not provide a complete representation of the Internet's structure. Further, increasing the number of probing stations quickly reaches a point of diminishing returns [13]. While these studies have provided

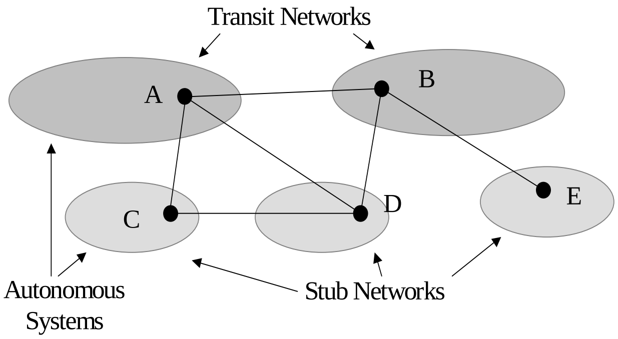


Figure 1: Example of Internet routing amongst separately administered, or autonomous, domains. A domain is classified as stub if the path connecting two nodes traverses the domain only if one of the nodes is in the domain. A domain is classified as transit if it does not have this restriction. The practice of having certain domains provide transit for network traffic, while others do not, imposes a hierarchical structure.

insightful techniques to improve the data collection process, thus far, no methods have been reported for fully resolving these issues.

In this study, we are interested in better quantifying the role current sources of network metrics play in revealing Internet structural attributes. Our intent is not to fully resolve current Internet mapping ambiguities, but rather to gain insights into when representations based on various empirical data are sound, versus when they may be misleading. The remainder of this paper is structured as follows. We begin with background information to introduce potential causes of artifacts in topology data. In Section 3, we present our data collection methodology, and provide an overview of the datasets used in our study. We detail the results of our analysis in Section 4, and offer conclusions in Section 5.

2. BACKGROUND

The Internet is a loose confederation of separate routing domains, or Autonomous Systems (AS). This structure enables routing within a domain and to adjacent domains to be independently administered. Routing within a given domain is achieved using Interior Gateway Protocols (IGP), such as OSPF and RIP. Routing between domains is achieved using Exterior Gateway Protocols (EGP) – BGP is the Internet's single de facto EGP [8]. Because our study is focused on the Internet's inter-domain topological attributes, we are most interested in the structure imposed by BGP routing policies. The two primary data collection methods used for inter-domain topology analysis are BGP routing tables and router-level probing. In this section, we discuss the major issues with each of these data sources.

For the remainder of this paper, we will refer to the network of interconnections between routers as a graph, $G(V,E)$, where V is the set of nodes (or routers) and $E \subseteq V \times V$ is the set of links, $deg(v)$ is the degree of node v. We use (u,v) to refer to the edge connecting nodes u and v, and Path(u, v) to refer to the sequence of edges traversed by a message flowing from node u to v.

2.1 BGP Routing Table Issues

A BGP router learns of routes to remote networks, specified by an IP address prefix, from its BGP peers. BGP peers are routers with which a given router maintains a BGP session. For each IP address prefix, a BGP router maintains a route entry, including the AS_PATH. The AS_PATH is intended to represent the sequence of AS's which would be traversed by a packet destined for the associated network if that packet were forwarded to the route entry's specified next hop router. Each entry of the AS_PATH is the uniquely assigned AS Number of the AS to be traversed. Therefore, each AS_PATH provides a list of inter-AS edges. As the following subsections illustrate, there are a number of issues with using this information for inter-domain topology analysis.

2.1.1 Partial Information

A router's BGP routing table provides complete information on the inter-AS edges that could be used to reach every known IP address from this router. The fact that the information is from a single viewpoint within the network is the source of the first set of issues with using BGP routing tables for topology analysis.

Single Viewpoint: While edges that radiate outward from the router's vantage point toward remote networks are well represented, edges between destination and intermediate domains generally are not. Consider the network depicted in Figure 1. The domains *A* and *B* represent networks that are providing transit services to *C*, *D* and *E*. Because *A* is directly connected to *C* and *D*, the router at A is unlikely to have edge (*C*,*D*) in its routing table.

Route Filtering: The absence of (*C*,*D*) in A's routing table may be due solely to the fact that it is not required to reach any other network, but this is unlikely due to the transit-stub relationship depicted in Figure 1. That is, even if link (*A*,*D*) were to fail, (*C*,*D*) is likely to still be absent from *A*'s routing table, due to route filtering. Routes that are intentionally not advertised to peers are said to be filtered. Filtering can be used to ensure traffic from the transit networks, such as *A* and *B*, are not routed through stub networks, such as *C* and *D*. Both *C* and *D* may choose to filter routes containing (*C*, *D*) to either *A* or *B*. To detect (*C*, *D*) in this scenario, a study would need to specifically include routing data collected from either *C* or *D*.

Route Selection: A router will maintain multiple route entries, or paths, for a given remote network, and denote one of the entries as the preferred, or "best," route. However, a router that receives multiple advertisements for paths to the same destination network may also be configured to discard route advertisements that are less desirable than entries representing alternate paths – this is typically to limit the size of the routing tables. If one or more edges of a discarded route are not in alternate paths maintained by the router, this edge will not be represented in the routing table.

2.1.2 Intra- versus Inter- domain routing

BGP routing tables represent only the information that is propagated to BGP. Much of the routing policy within an AS is not propagated to BGP routing tables. The separation of intra-domain routing protocols and inter-domain routing protocols can result in an `AS_PATH` that does not reflect how packets are actually forwarded. The following are examples of polices that may not be represented in BGP advertisements.

Static Routes: A router's BGP advertisements may not reflect how that router actually forwards packets. One example is when routes are statically defined within an AS, but are not propagated to the BGP routing tables. Such routes would not be represented in the BGP routing tables of the local AS or the routing tables of its neighbors.

Source Routing: The term source-routing is used to describe a policy in which a router forwards all packets with a given source IP address, to a specific next hop. For example, router *B* may be configured to forward all packets with a given source address to *A*. If *B* filters route entries received from *A*, packets would traverse (*B*, *A*) even though this link is not represented in the tables of routers adjacent to *B*.

Multi-hop BGP sessions: A pair of BGP routers may be logically, but not physically, adjacent. This configuration generally entails statically configuring the router pair to forward packets via an intermediate node. The router pair may then establish a TCP session via that intermediate node. Packets will be forwarded to the link connecting the intermediate node, and therefore the intermediate node would be represented in the `traceroute` results. However, the BGP advertisements generated at either of these nodes will not reflect the intermediate node.

2.2 Traceroute Issues

An alternative to basing topological analysis on BGP routing tables is to actively probe end-to-end paths at the router-level. Router-level probing techniques, such as `traceroute`, generally rely on eliciting ICMP messages from all the packet routers along a network path. We will use the small network depicted in Figure 2 throughout this section to illustrate the major issues with basing Internet structural analysis on data collected with `traceroute`.

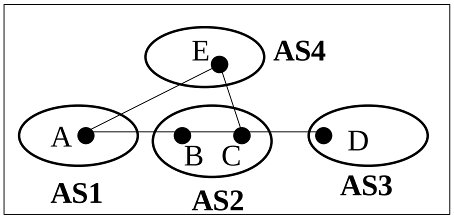


Figure 2: Example inter-domain network.

We begin by reviewing the operational characteristics of `traceroute`. A `traceroute` command issued at *A* to probe the path to *D* would begin by generating a UDP message with a Time-to-live (TTL) of 1. This message would be forwarded to *B. B* would detect the expiration of the TTL, send an ICMP Time Exceeded message to *A*, and discard the UDP packet. Because the source address of the IP packet containing the ICMP message is set to *B*'s address, `traceroute` is able to determine the first hop in the path from *A* to *D*. `Traceroute` increments the TTL and repeats the test until a response is received from *E*, or it times out waiting for a response. In the following subsections, we discuss issues with using `traceroute` results for topology analysis.

2.2.1 AS number resolution

The Internet's address space is allocated in a hierarchical manner. Blocks of the address space are allocated to Regional Internet Registries (RIRs), which, in turn, allocate address space to local registries who assign the addresses to end users [15]. Address space blocks are uniquely identified by IP address prefixes. Each AS is assigned a unique AS number, and an address space block. Each of the RIR's, ARIN, RIPE, and APNIC, maintain a database of the address space allocation for their region, and support database queries via the "whois" protocol. Therefore, the AS number of an IP address can be determined by querying the whois server of the appropriate regional server. Moreover, the Routing Arbiter Database [RADB] mirrors several regional databases, as well as other databases maintaining routing policy [16].

A commonly used version of `traceroute,` referred to as the NANOG-traceroute [9], maps the IP address received in ICMP probe reply to its corresponding AS number by querying the RADB. Specifically, it sends a query to whois.ra.net and parses the Origin field. An RADB response may list multiple Origins for a given query. The Origin with the most specific, i.e., longest, IP address prefix corresponding to the address in question prevails.

Registry Data: The method used to map the IP address in the ICMP message to an AS number represents the first issue. While some of the regional registries require network administrators to accurately maintain their registry data, others do not. Because changes are not automatically propagated to the registries, this information may be outdated, or simply incorrect.

Multiple AS numbers: A single service provider may register multiple AS numbers, but manage the corresponding address pools as a single routing domain. Suppose, for example, the administrator of the domain containing *B* and *C* in Figure 2 had registered 2 AS numbers, AS2 and AS2'. *B*'s interface to *A*, and *C*'s interface to *D* could be assigned addresses from AS2, while *B* and *C*'s internal interfaces could be assigned addresses from AS2'. A `traceroute` probe would indicate an AS path length of 4, whereas the BGP AS_PATH would indicate a length of 3.

2.2.2 ICMP message generation

Additional issues arise from differing implementations of ICMP message generation. The ICMP specification [10] does not state whether the source address of the ICMP reply should be the address of the interface on which the evoking packet was received, or the address of the interface on which the ICMP reply is sent. Referring again to Figure 2, suppose *C*'s interface to *E* is assigned from the address pool of the AS to which it connects, AS4. If *C* is configured to route packets to *A* via *E*, then a probe from *A* to *D* would indicate a path of AS1, AS2, AS4, AS3. Conversely, the BGP AS_PATH would correctly indicate a path of AS1, AS2, AS3.

As the issues described in this section illustrate, there are a number of difficulties in utilizing both BGP routing table data and end-to-end probing for Internet topology analysis. The goal of our study was not to develop or deploy new tools to address these issues. Instead, the intent was to quantify the effect of these issues on the resulting datasets, and to gain insights into which tools are more appropriate for which structural attributes. In the following section, we describe the methodology used to collect data used for this experiment.

3. EXPERIMENTAL METHODOLOGY

We collected four datasets representative of measurements typically used for Internet topology analysis. The goal for our first dataset, *D1*, was to collect AS path information representing the forward and reverse paths between a pair of nodes. We enlisted Looking Glass sites [3] distributed throughout the world for this purpose. Looking Glass sites provide an HTTP interface to invoke `traceroute` commands to specified IP addresses, and to query the site's local BGP router for the AS_PATH associated with an IP address. We identified 92 sites in which both the BGP query facility and the `traceroute` facility were active. Table 1 provides a list of the Looking Glass sites used in our study.

Looking Glass Site	Location	Looking Glass Site	Location
212.49.128.150	Spain	noc.tele.dk	Denmark
alice.eng.level3.net	San Jose	noc.tele.net	Switzerland
as5388.net	United Kingdom	noc.toplink.net	Herrenberg
bianaoh.cc.columbia.edu	New York	noc.urbanet.ch	Switzerland
dan.beesky.com	United States	noc.villageworld.com	New York
debby.sunrise.ch	Switzerland	noc.wsisiz.edu.pl	Poland
doom.net	Massachussetts	onet.on.ca	Ontario
euro-guest.genuity.net	Frankfurt	pegas.carrier.kiev.ua	Kiev
home.mobikom.net	Bulgaria	probe.global-one.nl	Netherlands
kix.net	Seoul	ptlduh00.eli.net	Virginia
lg.1anetworks.net	United Kingdom	reporter.teleglobe.net	Palo Alto
lg.above.net	Atlanta	spirit.interware.hu	Hungary
lg.broadwing.net	Virginia	spnet.net	Bulgaria
lg.carrier1.net	London	stat.cybercity.dk	Denmark
lg.citec.net	South Africa	statistics.kpnqwest.net	Washington, D.C.
lg.cohaesio.net	Denmark	stats.deine.net	Paris
lg.conxion.net	California	stats.mia.net	Wisconsin
lg.drift.telia.dk	Denmark	stats.solnet.ch	Switzerland
lg.dtr.fr	Lyon	support.bbc.co.uk	London
lg.euronet.nl	Amsterdam	traceroute.nnit.dk	Denmark
lg.hea.net	Dublin	widell.net	Sweden
lg.lan.switch.ch	Zurich	www.ams-ix.net	Amsterdam
lg.lasting.ro	Romania	www.bbeng.gxn.net	New York
lg.noc.netscalibur.de	Germany	www.demos.net	New York
looking-glass.in.bellnexxia.net	Toronto	www.doema.wirehub.nl	Netherlands
looking-glass.optus.net.au	Australia	www.ebone.net	Bratislava
looking-glass.taide.net	Sweden	www.ecs-ip.net	Amsterdam
looking-glass.teaser.fr	France	www.eng.nac.net	London
lookingglass.tops.net	Bonn	www.gigapop.gen.tx.us	Texas
mail.kamp.net	Frankfurt	www.gitoyen.net	France
neptune.dti.ad.jp	Japan	www.inetcomm.net	Russia
netcollect.kpn.net	Amsterdam	www.inoc.imnet.ad.jp	Tokyo
netcon.internet.fo	Torshavn	www.intelideas.net	Madrid
netmon.grnet.gr	Athens	www.macomnet.net	New Jersey
netstat.netone.com.tr	Istanbul	www.mediasat.ro	Romania
nms1.shinbiro.com	Korea	www.nat.bg	Bulgaria
noc.as8807.net	Denmark	www.noc.easynet.net	Delaware
noc.cetlink.net	Atlanta	www.noc.itgate.net	Milan
noc.colocall.net	Ukraine	www.nordu.net	Stockholm
noc.comstar.ru	Russia	www.opentransit.net	London
noc.datagrama.net	Spain	www.proxad.net	France
noc.kiev.sovam.com	Kiev	www.ripe.net	Amsterdam
noc.ngdc.net	Copenhagen	www.vianw.net	San Jose
noc.petrel.net	Switzerland	www.xmission.com	Utah
noc.runnet.ru	Russia	www.zimage.delbg.com	Bulgaria
noc.support.nl	Netherlands	www2.pt.lu	Luxembourg

Table 1: Looking Glass sites used. For those Looking Glass sites that allow queries via different sites, the location represents the site selected.

Each measurement consisted of a random pairing of two hosts with a `traceroute` command and a BGP query initiated in both directions. Measurements were made at Poisson intervals with a mean of 10 minutes between measurements initiated by the same host. Of the 116302 attempted measurements of 8464 uniquely paired hosts, 12% failed due to an error at one of the hosts or an error in reaching one of the hosts. An additional 21% were incomplete due to the server failing before it delivered its results, inability of the probe to elicit a response for each node along the path, or inability of the probe to reach the requested target in either or both directions.

Additionally, measurements that met any of the criteria listed in Table 2, or for which we were unable to capture a forward and reverse BGP path, as well as a forward and reverse `traceroute` in the same time period were discarded. We did not discard measurements with routing pathologies, such as a routing loop, as long as each of the paired probes reached their target. Using these techniques, we were able to collect 26978 fully paired `traceroute` and BGP queries representing 2840 unique route pairs.

The second dataset, *D2*, was collected from the Oregon Route Views website [2]. The website provides access to an archive of routing table snapshots for the Oregon Route View project's BGP router. The Oregon router maintains multi-hop BGP sessions with 57 other BGP routers. These peered routers are widely distributed over the Internet, both from a geographic and network-layer topological perspective, and represent the networks of most major existing Internet Service Providers. For each IP address prefix, the Oregon router maintains a list of all the AS paths it has learned from all of its peers. Therefore, retrieving the routing table of the Oregon router provides all of AS path data advertised by all of its BGP peers. The *D2* dataset is a collection of all of the BGP snapshots for the time period during which *D1* was collected. The snapshots were collected at 2-hour intervals.

Measurement Discard Criteria
Origin `traceroute` server not responding
Incomplete `traceroute` output
Node address in 10.x.x.x, 172.16.x.x-172.32.x.x, or 192.168.x.x range.
Route did note terminate in target AS
Intermediate node did not respond to ICMP echo
No matching reverse probe for same time period

Table 2: Criteria for Discarding Measurements.

Table 3 summarizes the characteristics of *D1* and *D2*. While the number of unique AS's encountered in *D1* is small in comparison to the remaining datasets, we argue that the *D1* measurements are representative of routes traversing the Internet core. Our argument is not formal, but is based on recent studies [4,5,11] indicating the Internet core comprises a relatively small number of AS's that provide transit for the majority of the remaining AS's, which are stub networks. We note that *D1* included 100% of the top 20 AS's in *D2* when ranked according to degree. Note that these top 20 nodes represent over 40% of the total edges discovered in *D2*.

Dataset	*D1*	*D2*	*D3*	*D4*
Date Collected	3/2002	3/2002	4/2002	4/2002
Collection Duration (days)	18	18	11	11
Data Source	LG	ORE	LG	ORE
Number of nodes/AS's	337	13054	7640	13226
Number of edges	1937	53816	25812	55410

Table 3: Dataset summary. LG represents the Looking Glass sites listed in Table 1 and ORE represents the Oregon Route View archive.

The Looking Glass and Oregon Route Views websites were also used to collect datasets *D3* and *D4*, respectively. *D4* was collected in the same manner as *D2*, except it represents the time period during which *D3* was collected. Unlike *D1* however, the goal of *D3* was not to collect forward and reverse paths between pairs of hosts. Instead, the goal was to capture path information for a large number of random destinations. We used the first routing table in *D4* to generate a list of IP address prefixes. Each measurement consisted of randomly selecting a Looking Glass server from the pool of servers listed in Table 1. A `traceroute` and BGP query were directed to this server. The target of the paired query was generated by randomly selecting an IP address prefix derived from *D4*, and then generating an IP host address with the corresponding prefix. While this meant that many of the probes were directed at an IP address for which the host did not actually exist, the valid IP prefix enabled the `traceroute` probe and BGP query to collect path information to the target network.

D3 measurements were also made at Poisson intervals with a mean of 10 minutes between measurements initiated by the same host. Of the 62645 attempted measurements, 14% failed due to an error in reaching or querying the Looking Glass host. An additional 23% were incomplete due to the server failing before it delivered its results, inability of the probe to elicit a response for each node along the path or inability of the probe to reach a host in the target AS. We were able to successfully collect `traceroute` and corresponding BGP path data for 27185 unique paths.

4. RESULTS

In the previous section, we collected datasets from a wide variety of vantage points within the Internet. By collecting both BGP routing tables and router-level probes, we have ensured the data represents both the advertised portion of routing policy, and the packet forwarding behavior of the corresponding paths. We have also pointed out a number of issues with current data sources for Internet topology analysis – issues from which our datasets also suffer. In this section, we attempt quantify the effect these measurement artifacts have on key topological attributes.

4.1 AS Path Asymmetry

We begin our evaluation with an analysis of a well-known property of Internet topologies – AS path asymmetry. Paxson [12] defined route asymmetry as the property of having Path(u,v)$\neq$Path(v,u) for any $u,v \in V$, and used `traceroute` probing to show that nearly half of all Internet paths included a major asymmetry. That is, about 20% of the end-to-end paths differed in two or more of the cities visited and about 30% differed in the AS's visited. Paxson's study was completed on data collected in 1995.

We used the *D1* dataset for this particular analysis. For those IP addresses that were not mapped to an AS number by the `traceroute` output, we resolved the AS number by querying the routing registry in a manner similar to the NANOG-traceroute described in Section 2.2. However, if the whois.ra.net query failed, the following order was used to query regional databases: whois.ripe.net, whois.apnic.net, whois.nic.mil, whois.arin.net. If multiple Origin records were reported for the longest corresponding IP address prefix, the most recent record was used. If the response included the AS Name assigned the IP address range of the address in question, but not the AS Number, we considered all addresses within the specified IP Address range to be in the same AS.

Not surprisingly, we found the AS path asymmetry based on data collected in 2002 was significantly higher than what had been reported for 1995 data. However, at 69.8% AS path asymmetry, the size of the increase was surprising. However, several factors may be artificially inflating this statistic, including ICMP message generation issues, outdated routing registry records, and multiple AS's managed as a single domain.

Our first target was the procedure for mapping IP addresses to AS numbers. To gain insights on this issue, we re-evaluated the asymmetry for *D1*, this time using *D2* to perform IP address to AS number resolution. Specifically, we retrieved the `AS_PATH` for the IP address's network prefix from *D2*, and used the final AS number in the `AS_PATH` as the AS number for the corresponding IP address. We refer to this method as BGP-AS-resolution. Using BGP-AS-resolution, we calculated AS asymmetry of only 61.4%. While BGP-AS-resolution is based on up-to-date information, it does not reduce potential inaccuracies due to multiple AS's managed as a single domain or ICMP message generation.

We investigated the ICMP message generation issue first. Recall from section 2.2, if the source IP address for the ICMP reply is set to the interface on which the ICMP message is sent and this interface's address was assigned from a neighboring AS's pool, `traceroute` would incorrectly indicate the neighboring AS was traversed. We began by checking the source code of the IP stack for AIX, FreeBSD and Linux. The FreeBSD and AIX implementations set the source IP address to the address of the interface on which the `traceroute` probe is received, and therefore would not suffer from incorrectly reporting the neighboring AS. We also tested the behavior on Cisco 7500 routers and Windows 2000 servers in our lab. Both the Cisco routers and the Windows 2000 system set the source IP address to the interface on which the traceroute probe was received. However, the Linux IP stack sets the IP source address to the interface on which the ICMP message is sent.

Without a mechanism to determine the IP implementation of intermediate routers, we were unable to isolate which routes erroneously included neighboring AS's. Therefore, our next test was to evaluate asymmetry without using `traceroute`. Specifically, we calculated asymmetry for the same routes using the BGP `AS_PATH` of *D1*. Because the BGP `AS_PATH` was collected from the router local to the Looking Glass host originating the corresponding `traceroute`, it should predict the AS path to be followed by the `traceroute` probe – with the exception of the anomalies listed in Section 2.1.

Many of the BGP routers tracked multiple `AS_PATH`'s to a given destination, and labeled one of the paths as "best." The "best" qualification may be assigned based on attributes such as AS path length or configured policy. We used only the BGP path labeled "best" in calculating AS path asymmetry.

When calculated using `AS_PATH`, the asymmetry for the same route set was 56.3%. Figure 3 compares the hop difference distribution for each of the 3 data sources. While the distribution of routes with 1 or more hop differences was not significantly different for each of the methods, the nearly 15% difference in number of fully symmetric paths under BGP `AS_PATH`, as opposed to `traceroute` with RIR AS resolution, clearly indicates a difference in the `traceroute` path and that predicted by `AS_PATH`. Our next challenge was to better quantify this difference on a per route basis.

4.2 BGP AS_PATH prediction of `traceroute` AS path

In this section, we compare the AS path reported by `traceroute` with that predicted by the BGP `AS_PATH` of the Looking Glass router local to the node performing the `traceroute`. We started with a relatively simple attribute of the Internet's AS topology – AS path length. For this particular test, we used the *D3* dataset so paths would not be limited to those between Looking Glass sites. As described in Section 4.1, we resolved IP addresses to AS numbers by using the Oregon router's BGP tables for the corresponding time period (*D4*). We discarded those routes that did not reach at least one node in the target AS.

We found the average AS path length was 4.49 when calculated using `traceroute` probe data, and 4.15 when calculated using BGP `AS_PATH` for the corresponding measurement. Likewise, the path length distributions, plotted in Figure 4, showed little difference

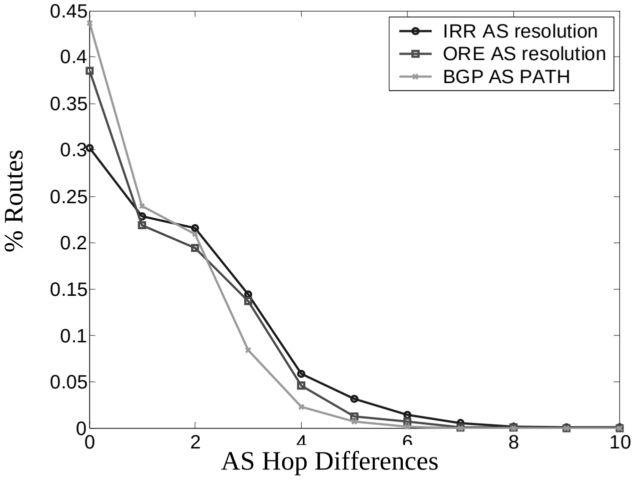


Figure 3: Comparison of AS asymmetry when calculated using `traceroute` with IRR resolution of IP addresses to AS numbers, with Oregon Route View resolution of AS numbers, and using BGP AS_PATH data.

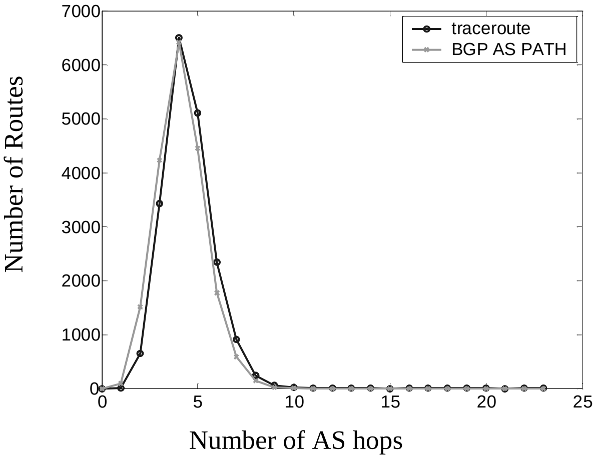


Figure 4: Comparison of AS path length distribution for traceroute data and BGP AS_PATH data.

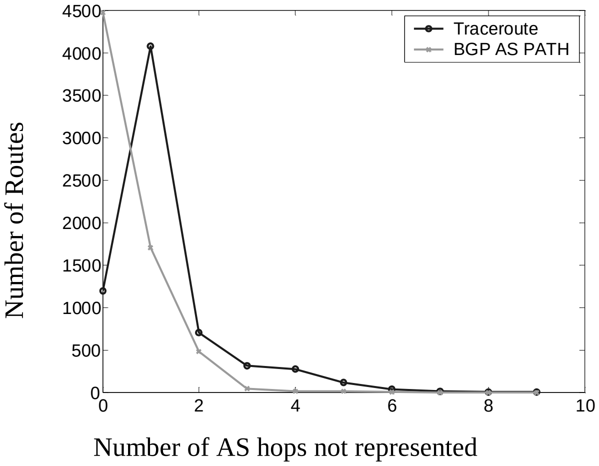


Figure 5: Comparison of AS hops not represented the corresponding traceroute/BGP AS path. Only paths in which the traceroute path did not match the BGP AS_PATH are represented.

when calculated with either traceroute, or corresponding BGP AS_PATH data.

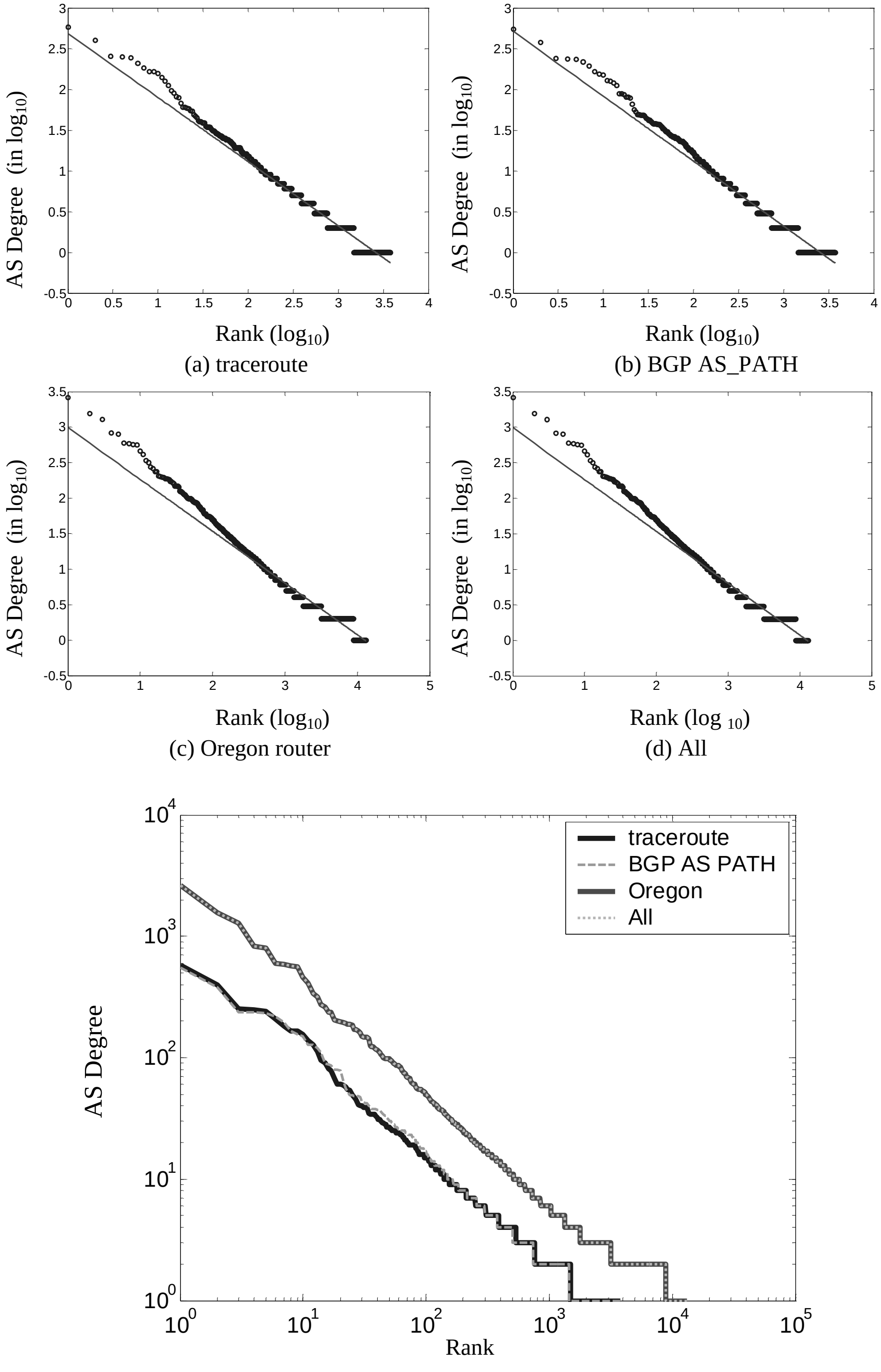
However, when the data was compared on a per-route basis, the differences were significant. For example, we found that 32.7% of all measured traceroute paths indicated a different path length than the corresponding BGP AS_PATH. Recall from section 2.1, issues with ICMP message generation can cause traceroute paths to incorrectly include AS's not actually traversed – thus inflating path length when compared to the corresponding BGP AS_PATH. Similarly, multi-hop BGP sessions can cause BGP AS_PATHs to incorrectly omit AS's actually traversed – also causing traceroute path lengths to appear inflated. Static routes and source-routing not propagated to BGP are likely to result in AS_PATHs which also differ from actual forwarding behavior – in which case, the traceroute path lengths are not strictly longer, or shorter, than the corresponding AS_PATH.

Figure 5 plots the distribution of routes according to the number of AS's which are not represented in the corresponding traceroute or BGP AS_PATH. For example, there were approximately 500 routes in which the BGP AS_PATH included 2 AS's which were not represented in the traceroute results, and there were approximately 700 routes in which the traceroute results included 2 AS's which were not in the BGP AS_PATH. The format of this graph was chosen to illustrate that the traceroute paths were not strictly longer, or shorter, than the AS_PATH's, and vice versa. Additionally, the number of traceroute paths that exhibited 2 or more nodes not in the corresponding BGP path was similar to the number of BGP paths exhibiting the reverse. The story for 0 or 1 hops, however, was quite different. More precisely, 74% of the routes with different path lengths were different because the traceroute path included a single additional node that was not included in the BGP AS_PATH, and the corresponding BGP AS_PATH contained no nodes that were not represented in the traceroute path.

4.3 AS Degree

A number of recent studies have focused on characterizing the Internet's inter-domain topology according to the distribution of edges per AS node, or degree. Representing the Internet's degree distribution with a purely mathematical formulation, as opposed to routing policies and hierarchies, would significantly simplify network analysis. In this section, we use the *D3* and *D4* datasets, along with the insights gained from the asymmetry and path prediction experiments, to investigate differences in characterizing AS degree.

We calculated the AS degree for nodes represented in *D3*'s traceroute results, *D3*'s BGP AS_PATH results, and *D4*'s Oregon routing tables. For each of these datasets, we sorted the AS's in descending order of AS degree and plotted their distribution in Figure 6 (a), (b), and (c), respectively. We also created a third dataset, labeled "All" in Figure 6(d), to represent the aggregate of all nodes and edges discovered by all 3 methods. Each plot includes a best



(a) traceroute

(b) BGP AS_PATH

(c) Oregon router

(d) All

(e) Comparison of distributions.

Figure 6: AS edge degree (y-axis) versus rank (x-axis) in sequence of decreasing degree. Graphs (a), (b) and (c) represent the degree distributions based on the traceroute and BGP AS_PATH of *D3* and the Oregon routing tables of *D4*, respectively. The distribution in graph d represents the aggregation of edges discovered in any of these three datasets. (e) provides a side-by-side comparison. The solid line in (a)-(e) represents the best exponential fit for the distribution.

exponential fit for the distribution. The correlation coefficient for the "traceroute," "AS_PATH," "Oregon," and "All" distributions and their corresponding best exponential fit were 0.97, 0.97, 0.96, and 0.96, respectively. These high correlations were achieved without discarding any outliers.

The comparison in Figure 6(e) makes it easier to see several important features of the distributions. First, the distribution obtained from *D3*'s traceroute data and BGP AS_PATH data are almost completely overlapped. The number of nodes discovered with traceroute was slightly more than discovered with AS_PATH – traceroute data included 18 more nodes than the AS_PATH data. This supported earlier findings that, of the nodes that differed, the majority differed because the traceroute path included a single additional node. It was, at first, surprising that the BGP data included nearly 200 edges that were not represented in the traceroute data. We did a visual inspection of many of the routes that produced edges that were not represented in traceroute. In those routes we inspected, the traceroute included a single additional node, and that node was the AS of an exchange point. Moreover, it was the edge between the AS's directly before and after the exchange point in the traceroute that was represented in the BGP data and not the traceroute data. We were not able to validate whether this anomaly was caused by a multi-hop BGP session traversing the exchange point, or the ICMP message generation issue.

Another interesting observation regarding the distribution for `traceroute` and `AS_PATH` data, is that it appears to flatten slightly where the highest degree edges are represented, and taper slightly where the lowest degree edges are represented. This is not surprising and, in comparison with the Oregon data, indicates not all of the stub nodes were discovered. It is reasonable to assume that discovering these additional stub nodes would lead to discovering additional edges for the highest degee nodes.

It is also interesting to note that the distributions for the Oregon data, and the aggregation of all three sources were almost completely overlapped. That is, of the more than 3700 nodes discovered in the `traceroute` and BGP `AS_PATH` data, only 35 were not already represented in the Oregon data.

5. CONCLUSIONS

There has been a recent flurry of research into the structural attributes of the Internet. Analyses are generally based on BGP routing tables and `traceroute`-like probing. In addition to pointing out a number of issues with current data sources for Internet topology analysis, we have reported on the effect these issues have on key topological attributes. Our analysis is based on data collected from a wide variety of vantage points within the Internet. By collecting both BGP routing tables and router-level probes, we have ensured the data represents both the advertised portion of routing policy, and the packet forwarding behavior of the corresponding paths. Our experiments have demonstrated that the effect on certain attributes, such as the average path length and the AS degree distribution can be minimized through careful data collection from a number of network vantage points. We have also shown that using this same data to model other attributes, such as the actual forwarding path between a pair of nodes, or the level of AS path asymmetry, can be very misleading.

ACKNOWLEDGEMENTS

We thank Alan Crosswell and Bartlomiej Solarz-Niesluchowski for valuable discussions on router configuration and Internet Service Provider routing policies. We also thank Jisoo Lee for configuring the Cisco routers used in the ICMP message handling experiments.

REFERENCES

1. Cooperative Association for Internet Data Analysis (CAIDA). http://www.caida.org.
2. D. Meyer, University of Oregon Route Views Project. http://www.antc.uoregon.edu/route-views.
3. T. Kernen, Traceroute Organization, http://www.traceroute.org.
4. M. Faloutsos, P. Faloutsos, C. Faloutsos, “On the Power-Law Relationships of the Internet Topology,” Proceedings of ACM SIGCOMM, Sept. 1999.
5. Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, “The Origin of Power Laws in Internet Topologies Revisited,” http://topology.eecs.umich.edu/archive/origin.ps.
6. Y. Rekhter, T. Li, “A Border Gateway Protocol 4 (BGP-4),” Internet Engineering Task Force Request for Comment 1771, March 1995.
7. V. Jacobson, Traceroute software, ftp://ftp.ee.lbl.gov/traceroute.tar.gz, 1989.
8. J. Stewart, *BGP4: Inter-Domain Routing in the Internet*, Addison-Wesley, 1998.
9. NANOG-traceroute. ftp://ftp.login.com/pub/software/traceroute.
10. J. Postel, “Internet Control Message Protocol,” Internet Engineering Task Force Request for Comment 792, September, 1981.
11. K. Calvert, M. Doar, and E. W. Zegura, "Modeling internet topology, " *IEEE Communications Magazine*, June 1997.
12. V. Paxson, “End-to-End Routing Behavior in the Internet,” *IEEE/ACM Transactions on Networking*, Vol.5, No.5, pp.601-615, October 1997.
13. P. Barford, A. Bestavros, J. Byers, M. Crovella, “On the Marginal Utility of Network Topology Measurements,” ACM SIGCOMM Internet Measurement Workshop, November 2001.

14. A. Broido, K. Claffy, “Internet Topology: connectivity of IP graphs,” SPIE International Symposium on Convergence of IT and Communication, August 2001.
15. D. Karrenberg, G. Ross, P. Wilson, L. Nobile, "Development of the Regional Internet Registry System," The Internet Protocol Journal, http://www.cisco.com/warp/public/759/ipj_4-4/ipj_4-4_regional.html.
16. Routing Arbiter Database. http://www.radb.net.