Towards an Accurate AS-Level Traceroute Tool

Zhuoqing Morley Mao
UC Berkeley
zmao@cs.berkeley.edu

Jennifer Rexford
AT&T Labs–Research
jrex@research.att.com

Jia Wang
AT&T Labs–Research
jiawang@research.att.com

Randy H. Katz
UC Berkeley
randy@cs.berkeley.edu

ABSTRACT

Traceroute is widely used to detect routing problems, characterize end-to-end paths, and discover the Internet topology. Providing an accurate list of the Autonomous Systems (ASes) along the forwarding path would make traceroute even more valuable to researchers and network operators. However, conventional approaches to mapping traceroute hops to AS numbers are not accurate enough. Address registries are often incomplete and out-of-date. BGP routing tables provide a better IP-to-AS mapping, though this approach has significant limitations as well. Based on our extensive measurements, about 10% of the traceroute paths have one or more hops that do not map to a unique AS number, and around 15% of the traceroute AS paths have an AS loop. In addition, some traceroute AS paths have extra or missing AS hops due to Internet eXchange Points, sibling ASes managed by the same institution, and ASes that do not advertise routes to their infrastructure. Using the BGP tables as a starting point, we propose techniques for improving the IP-to-AS mapping as an important step toward an AS-level traceroute tool. Our algorithms draw on analysis of traceroute probes, reverse DNS lookups, BGP routing tables, and BGP update messages collected from multiple locations. We also discuss how the improved IP-to-AS mapping allows us to home in on cases where the BGP and traceroute AS paths differ for legitimate reasons.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network monitoring*

General Terms

Measurement, Management

Keywords

Network measurements, AS-level path, Internet topology, Border Gateway Protocol

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SIGCOMM'03, August 25–29, 2003, Karlsruhe, Germany.
Copyright 2003 ACM 1-58113-735-4/03/0008 ...$5.00.

1. INTRODUCTION

Network operators and researchers would benefit greatly from an accurate tool for reporting the sequence of Autonomous Systems (ASes) along the path to a destination host. Designing a useful "AS-level traceroute" tool depends on having an accurate way to map the IP addresses of network equipment to the administering ASes. This problem is surprisingly difficult and existing approaches have major limitations, due to the operational realities of today's Internet. We propose a way to improve the IP-to-AS mapping of the infrastructure by comparing traceroute and BGP (Border Gateway Protocol) paths collected from multiple vantage points. This improved IP-to-AS mapping can be used as seed input for a tool that maps traceroute output to an AS-level path.

1.1 Motivation for AS Traceroute

Traceroute [1] is widely used to detect and diagnose routing problems, characterize end-to-end paths through the Internet, and discover the underlying network topology. Traceroute identifies the interfaces on a forwarding path and reports round-trip time statistics for each hop along the way. Despite its many well-documented limitations, traceroute is the only effective way to determine how packets flow through the Internet without real-time access to proprietary routing data from each domain. The tool is invaluable for network operators in identifying forwarding loops, blackholes, routing changes, unexpected paths through the Internet, and, in some cases, the main components of end-to-end latency. Researchers rely heavily on traceroute to study routing protocol behavior [2], network performance [3], and the Internet topology [4, 5, 6, 7].

In addition to the IP forwarding path, operators often need to know which ASes are traversed en route to the destination. Upon detecting a routing or performance problem, operators need to identify (and notify!) the responsible parties—often their compatriots in other ASes. This is a crucial part of diagnosing and fixing problems that stem from misconfiguration of the routing protocols [8] or serious equipment failures. For example, suppose that customers complain that they cannot reach a particular Web site. The operator could launch traceroute probes toward the destination and determine that a forwarding loop is to blame. However, correcting the problem requires a way to determine which AS (or set of ASes) has routers forwarding packets in the loop. Inaccurate information leads to delays in identifying and correcting the problem.

Researchers use the AS path information to construct AS-level views of the Internet topology [9] and to study the properties of the AS paths traversing this graph [10]. These AS-level "outputs" are used as "inputs" to research in a variety of areas, such as the placement and selection of Web content replicas. The accuracy of these studies hinges on having a sound way to determine the sequence of

ASes on a forwarding path. However, a recent paper [11] demonstrated that the various techniques for identifying the AS-level forwarding path lead to very different results for basic properties of the Internet topology, including path asymmetry and node degree. Having a good AS-level traceroute tool in the research community would make these studies more accurate. In addition, new research could focus directly on the properties of the AS-level forwarding path, such as identifying the ASes most responsible for forwarding anomalies and performance problems.

1.2 Difficulty of the Problem

Identifying the ASes along the forwarding path is surprisingly difficult. Traceroute infers the path by transmitting a sequence of TTL-limited packets and extracting the interface IP addresses from ICMP responses sent by the hops along the way. However, some hops do not return ICMP replies, and successive TTL-limited packets do not necessarily follow the same forwarding path. Mapping the IP-level hops to AS numbers is complicated and existing approaches have major limitations:

BGP AS path: A seemingly natural way to determine the AS path is to observe the routes learned via BGP, the interdomain routing protocol for the Internet. However, timely access to BGP data is not always possible from the vantage point of interest. Perhaps more importantly, BGP provides the *signaling* path (the list of ASes that propagated the BGP update message), which is not necessarily the same as the *forwarding* path (the list of ASes traversed by data packets). Although the two AS paths usually match, they may differ for various reasons such as route aggregation/filtering and routing anomalies [12]. In fact, the two paths may differ *precisely* when operators most need accurate data to diagnose a problem.

Internet route registry: Instead, the ASes in the forwarding path can be derived directly from the traceroute data by associating each traceroute hop with an AS number. The popular "NANOG traceroute" [13] and prtraceroute [14] tools perform *whois* queries to map each interface to an address block allocated to a particular AS. However, *whois* data are often out-of-date or incomplete, since institutions do not necessarily update the database after acquisitions, mergers, and break-ups, or after allocating portions of their address blocks to customers.

Origin AS in BGP routes: A more accurate and complete IP-to-AS mapping can be constructed from BGP routing tables by inspecting the last AS (the "origin AS") in the AS path for each prefix [15]. However, some traceroute hops map to multiple origin ASes (MOAS) [16] or do not appear in the BGP tables. The notion of "origin AS" blurs the many reasons why ASes introduce prefixes into BGP. In addition to originating routes for its own infrastructure, an AS may inject routes on behalf of statically-routed customers. Some ASes do not advertise their infrastructure addresses and others may announce the addresses of shared equipment at boundary points between domains. As a result, some traceroute AS paths appear to have AS loops, or extra or missing hops relative to the corresponding BGP paths.

In this paper, we identify the root causes of the differences between the traceroute and BGP AS paths, and propose techniques for identifying the "real" AS-level forwarding path.

1.3 Our Approach to the Problem

In practice, the signaling and forwarding AS paths do not always agree, due to route aggregation and forwarding anomalies. However, we believe that most discrepancies between the BGP and traceroute AS paths stem from inaccuracies in the IP-to-AS mapping applied to the traceroute data. We propose to improve this mapping by comparing BGP and traceroute paths from multiple vantage points. Our algorithms analyze measurement data to identify cases where a single "explanation" would account for the differences between many pairs of BGP and traceroute AS paths. These explanations build on an understanding of common operational practices, such as the presence of Internet eXchange Points (IXPs), where multiple ASes connect to exchange BGP routes and data traffic. The results of our algorithms are used to tune an initial IP-to-AS mapping derived from the BGP routing tables. We envision this as a continuous process where traceroute and BGP data are collected from many vantage points and used to compute an accurate IP-to-AS mapping as it changes over time. An AS traceroute tool running on end hosts would periodically download the latest IP-to-AS mapping and use it to compute and display the AS path associated with each traceroute probe the user launches. The paper makes five main contributions toward this end:

Measurement methodology: Our techniques depend on collecting traceroute probes, BGP update messages, BGP routing tables, and reverse DNS lookups, as discussed in Section 2.

Traceroute analysis: In Section 3, we analyze traceroute and BGP paths from eight locations, and construct an initial IP-to-AS mapping from BGP routing tables. Then, we present an initial comparison of the BGP and traceroute AS paths.

Resolving incomplete paths: Section 4 presents three simple techniques for resolving most traceroute hops that do not map to an AS number. We also introduce our approach of using internal router configuration data for checking our results.

Improved IP-to-AS mapping: Many mismatches between BGP and traceroute paths can be explained by IXPs, sibling ASes managed by the same institution, and ASes that do not advertise routes to their equipment. Section 5 proposes techniques that identify and "fix" some of these cases.

Legitimate mismatches: The traceroute and BGP AS paths may differ for valid reasons such as route aggregation, interface numbering at AS boundaries, the choice of source address in ICMP, and routing anomalies. Section 6 discusses how these factors may explain some of the remaining differences between the traceroute and BGP AS paths.

Validating our techniques is difficult without knowing the actual AS-level forwarding paths. Where possible, we compare results with publicly-available data, such as *whois* data and lists of known IXPs. We conclude in Section 7 with a summary of our contributions and a discussion of ongoing work.

1.4 Related Work

Recent measurement studies have quantified the differences between BGP and traceroute AS paths. The analysis in [11] showed that these differences have a significant impact on the characterization of the Internet topology. In parallel with our paper, the work in [17] used publicly-available data (such as *whois*, lists of known IXPs, and other Web sites) to test the hypothesis that many of the mismatches stem from IXPs and siblings; in contrast, our paper proposes heuristics for *identifying* IXPs, siblings, and other causes of mismatches to improve the IP-to-AS mapping. To improve the accuracy of AS graphs derived from traceroute, the work in [18] proposed techniques that identify border routers between ASes to correct mistaken AS mappings; this is an alternate approach that handles some of the inaccuracy introduced by IP-to-AS mappings derived from BGP tables. Traceroute data have been used in other studies that measure router-level topologies and map routers to ASes [4, 6]. Except for handling certain traceroute anomalies such as unmapped IP address, these studies did not focus on improving the accuracy of the IP-to-AS mapping derived from the BGP routing tables. Focusing solely on BGP AS paths, the work in [19, 20] presented algorithms for inferring AS-level commercial relationships, including siblings; however, these studies did not consider the influence of sibling ASes on the accuracy of traceroute AS paths.

In contrast to previous work, our paper focuses on automated techniques for improving the IP-to-AS mapping applied to the traceroute paths. Although we use publicly-available information for validation purposes, the techniques we propose do not depend on the availability of such data. Our work capitalizes on traceroute paths and BGP updates collected from multiple vantage points to a large number of destinations throughout the Internet. The techniques we apply to pre-process the measurement data limit possible inaccuracies from transient routing changes and unmapped hops in the traceroute paths. Our algorithms for identifying IXPs, siblings, and unannounced infrastructure addresses allow us to produce a more accurate estimate of the AS-level forwarding path from the raw traceroute data. This, in turn, enables us to focus our attention on the legitimate mismatches between the AS-level signaling and forwarding paths.

2. MEASUREMENT METHODOLOGY

This section presents our methodology for collecting traceroute and BGP paths from multiple vantage points, as shown in Figure 1. We select candidate prefixes and ultimately individual IP addresses to cover the routable address space. For each prefix we measure the forwarding path with traceroute and extract the BGP AS path from the routing table of the border router. We discard data for cases where the BGP AS path cannot be meaningfully compared with the traceroute path. We compute an AS-level traceroute path by mapping traceroute hops to AS numbers using the origin ASes extracted from a large set of BGP routing tables.

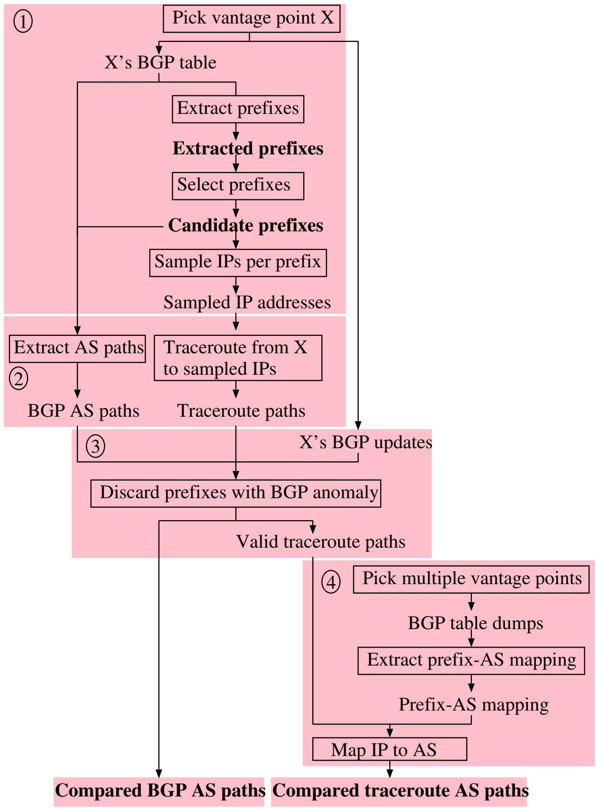


Figure 1: BGP and traceroute data collection.

2.1 Selecting Candidate IP Addresses

Starting with a list of routing table entries, we first identify the prefixes that cover the routable address space and then select two IP addresses within each prefix for traceroute probing.

Select prefixes: Ideally, we would like to learn the forwarding path to each live destination address from each vantage point. However, identifying all live IP addresses is challenging and sending traceroute probes to each destination would be prohibitively expensive. Instead, we select a set of prefixes that cover the routable address space to sample a wide range of forwarding and signaling paths. For each vantage point, we extract a list of prefixes from the BGP routing table of the (single) border router that connects this site to the Internet. However, some prefixes are never used to route traffic because of more specific subnets in the routing table. For example, no packet would use the 192.0.2.0/23 route if nested entries for 192.0.2.0/24 and 192.0.3.0/24 were available, due to the longest-prefix match forwarding paradigm. Other prefixes may be partially covered by subnets. For example, a table with routes for 8.0.0.0/8 and 8.128.0.0/10 would only use the 8.0.0.0/8 routing entry for destinations in 8.0.0.0/9 or 8.192.0.0/10; all other addresses in 8.0.0.0/8 would match 8.128.0.0/10. To identify these cases, we sort the list of prefixes based on the numerical values and mask length; this ensures that each prefix is followed immediately by all of its subnets. For each prefix, we identify the portion of the address space that would match this routing table entry and represent it as a list of address blocks. The algorithm runs in $O(n^2)$ time in the worst case that all n routing entries are subnets of a single prefix. In that case, the difference between a given prefix and all other prefixes preceding it in the sorted order need to be calculated. Prefixes like 192.0.2.0/23 that are covered by their subnets do not correspond to any portion of the address space, and are excluded from the list of candidate prefixes.

Sample IP addresses: Each candidate prefix has one or more IP addresses that match the routing entry using longest prefix matching. We select two IP addresses for each prefix for the sake of comparison; this is especially useful for studying the effects of route aggregation and filtering. Limiting ourselves to two addresses reduces the time required to collect the data. For each prefix, we arbitrarily select the first address block in the representation computed by our algorithm (*e.g.,* 8.0.0.0/9 for $\{$8.0.0.0/9, 8.192.0.0/10$\}$). We select the two IP addresses from the beginning and the middle of the block. That is, for a block with address Q and mask length N, we select IP addresses "$Q + 1$" and "$Q + 2^{32-N-1} + 1$" (*e.g.,* 8.0.0.1 and 8.64.0.1). Note that the addresses do not necessarily correspond to live hosts; some may be unused or assigned to parts of the infrastructure.

2.2 Obtaining Traceroute and BGP Paths

After selecting the IP addresses, we obtain both the traceroute and BGP paths from each vantage point.

Collect traceroute paths: We configure the traceroute software to send a single UDP packet for each TTL value and wait two seconds for an ICMP reply. In lieu of sending multiple packets per TTL value, we modified the traceroute source code to send a second packet only if the first attempt did not produce an ICMP response; for the second packet, we wait five seconds for a reply. To further reduce the overhead, we apply the modified traceroute with DNS resolution disabled; after completing the traceroute experiments, we perform a reverse DNS lookup for each unique IP address that appears in the traceroute output. For each destination address, we record a timestamp and the traceroute output.

Extract BGP AS paths: For each candidate prefix, we extract the corresponding AS path from the most recent (daily) dump of the local BGP table that occurred before the traceroute. If the table has no BGP route for the prefix (say, due to a BGP withdrawal since the initial table dump), we extract the AS path of the longest matching prefix. We preprocess the BGP AS path to collapse consecutive repeating ASes (*e.g.,* converting "701 88 88" to "701 88") that stem from AS prepending.

2.3 Discarding Based on BGP Properties

In some cases, comparisons between the BGP and traceroute paths are not meaningful, and we discard both paths:

BGP routing changes: Routing changes introduce uncertainty in the local BGP AS path during the period of a traceroute experiment. Starting with the most recent BGP table dump, we apply the sequence of update messages to track changes in the BGP AS path for each prefix over time. After a traceroute completes, we identify the most recent BGP route update for that prefix and use this AS path in our subsequent analysis. We also inspect a window of time before and after the traceroute for update messages for the prefix. If a BGP routing change occurs during this window, we exclude this prefix from our analysis. In our study, we apply a 30-minute window before and after each traceroute to account for delays in BGP routing convergence [21]. Still, we cannot ensure that the *forwarding* path remains the same throughout the traceroute experiment. For example, the forwarding path may fluctuate due to an *intra*domain routing change. In addition, some downstream AS might experience a BGP routing change for some subnet of the prefix that is not seen at local collection point. We can only ensure that the BGP AS path seen at our collection point is stable during the traceroute experiment.

Null AS paths: Each BGP table has a few routes with a null AS path. These routes correspond to prefixes belonging to the institution where we collected the BGP and traceroute data.

Private AS numbers: Some BGP AS paths contain private AS numbers in the range of 64512–65535. This can arise when a customer (using a private AS number) mistakenly leaks BGP routes learned from one upstream provider to another.

Apparent AS loops: BGP has a loop-detection mechanism where a router discards a route when its own AS number appears in the AS path. However, apparent AS-level loops can arise if a router is (mis)configured to prepend an arbitrary AS number that already appears elsewhere in the AS path.

AS_SET: In the usual case, the AS path information is encoded as a sequence of ASes. However, occasionally, when a router aggregates multiple BGP routes, the resulting AS path may include an *unordered* set of ASes from the original paths (in order to prevent loops). This makes it impossible to determine the sequence of ASes in the path(s).

2.4 Computing Traceroute AS Paths

Computing the AS-level traceroute path requires mapping the IP addresses in the path to AS numbers. We construct an initial mapping by combining BGP routing tables from multiple locations and extracting the last hop in the AS path (the "origin AS") for each entry. An individual IP address is mapped to the longest matching prefix. If a prefix has routes with multiple origin ASes (MOAS), we map the IP address to the *group* of ASes. After mapping each traceroute hop to an AS (or group of ASes), we collapse hops with the same mapping to produce the AS-level traceroute path.

3. TRACEROUTE ANALYSIS

In this section, we apply our measurement methodology to traceroute and BGP routing data collected from eight sites. We analyze the diverse ways the traceroute experiments can *end* and explain how we preprocess the data. We quantify the limitations of using Internet routing registry data to map the IP addresses to AS numbers and evaluate our approach of using BGP routing tables collected from multiple vantage points. Still, many of the hops in the traceroute paths have IP addresses that map to multiple ASes or do not appear in the BGP table.

3.1 Collecting Traceroute and BGP Updates

We collected detailed routing data from eight locations in the North America, as summarized in Table 1. The sites were chosen based on their topological diversity and our ability to collect both traceroute and BGP update data. At each location, we ran traceroute on machines one or more hops behind a single border router; the traceroute data were preprocessed to remove the initial hops between the probe machine and the border router. In AS 6431 and AS 25 we had root access to a Linux machine that ran our modified traceroute software to send a second TTL-limited probe upon receiving a "*" response from an intermediate hop; sending a second packet resulted in a successful ICMP reply in 7% of the cases. In other locations, we used a standard traceroute configured to send a single probe for each hop to reduce overhead and delay. At each site, we collected BGP updates in MRT format through a BGP session with the border router, along with daily dumps of the BGP routing table. At each location, the machines sending traceroute probes and logging the BGP updates had their clocks synchronized using NTP. For brevity, we present the results from the first three locations only; the results from other locations are similar.

Despite the topological diversity of our measurement points, our analysis would benefit from a larger number of data sets from different countries. On the surface, using the publicly-available traceroute servers would seem like a natural solution to this problem. However, BGP update messages are not available from these servers, although some support querying of the BGP routing table; this would have allowed us to poll a prefix's BGP route a few minutes before and after each traceroute experiment, in the hope of catching relevant BGP routing changes. However, the public traceroute servers typically impose a rate limit on requests issued from the same host, making it difficult to probe a large number of addresses in a reasonable amount of time. The long delay could span significant changes in the Internet topology and in the mapping of prefixes to ASes. In addition, the GUIs at the public servers typically do not support changes to traceroute parameters (*e.g.,* number of probes per hop, timeout for ICMP replies, and disabling DNS resolution). As such, although our methodology can be applied to an arbitrary number of vantage points, the analysis in this paper focuses on a smaller number of data-collection points under our direct control; where relevant, we comment on how the limited vantage points may affect our results.

Our analysis focuses on one set of traceroute experiments from each location; results from other dates produced very similar results. The eight traceroute data sets were collected between May and June in 2003. Table 2 reports the number of prefixes *extracted* from the local BGP routing table at the first three sites, following the steps outlined earlier in Figure 1. The table also lists the number of *candidate* prefixes used in the traceroute experiments (with two destination addresses per prefix), after applying the algorithm in Section 2.1; the other 1.3–1.4% of the BGP prefixes were not the longest matching route entry for any destination addresses. The *compared* prefixes excludes the cases where comparisons with the

Organization	Location	Dates in 2003	Upstream Provider (AS Number)
AT&T Research (AS 6431)	NJ, USA	June 6-9	UUNET (701), AT&T (7018)
UC Berkeley (AS 25)	CA, USA	June 6-8	Qwest (209), Level 3 (3356)
PSG home network (AS 3130)	WA, USA	April 30 - May 8	Sprint (1239), Verio (2914)
Univ of Washington (AS 73)	WA, USA	June 4-8	Verio (2914), Cable & Wireless (3561)
ArosNet (AS 6521)	UT, USA	May 1-6	UUNET (701)
Nortel (AS 14177)	ON, Canada	May 1-6	AT&T Canada (15290)
Vineyard.NET (AS 10781)	MA, USA	June 4-9	UUNET (701), Sprint (1239), Level 3 (3356)
Peak Web Hosting (AS 22208)	CA, USA	May 1-8	Level 3 (3356), Global Crossing (3549), Teleglobe (6453)

Table 1: Traceroute probing locations

	AS 6431	AS 25	AS 3130
Extracted	121259	124295	120996
Candidate	119550	122487	119340
Compared	118345	112120	117195

Table 2: Number of prefixes in the three datasets

	AS 6431	AS 25	AS 3130
Routing changes	0.3802%	0.5809%	0.3105%
Null AS paths	0.0058%	0.0064%	0.0000%
Private ASes	0.0000%	0.0000%	0.0008%
AS loops	0.0000%	0.0000%	0.0155%
AS_SET	0.0214%	0.0233%	0.0248%

Table 3: Prefixes excluded due to BGP properties

	AS 6431	AS 25	AS 3130
Expected	11.21%	11.24%	11.21%
Unexpected	14.37%	14.17%	15.00%
"*"	54.79%	55.48%	53.92%
"!"	12.15%	12.09%	12.48%
30 hops	7.47%	7.02%	7.40%

Table 4: Ending of the traceroute experiments

BGP AS paths were not meaningful. Table 3 presents a more detailed breakdown of the five cases, which account for less than 1% of the prefixes probed in the traceroute experiments. The rest of the candidate prefixes that are not compared are due to BGP table changes causing some long prefixes to disappear and failed traceroutes caused by routing problems. The compared prefixes form the basis of the analysis in the remainder of the paper.

3.2 Characterizing the Traceroute Results

Ideally, traceroute returns a complete list of IP addresses up to and including the destination. This requires each hop to return an ICMP TIME_EXCEEDED message with the address of the corresponding interface and the destination host to return a PORT_UNREACHABLE message. In practice, the traceroute paths *end* in five different ways, as summarized in Table 4:

Expected final address: Only around 11% of the paths end with a PORT_UNREACHABLE message from the target IP address. In a way, this is not surprising because the destination address does not necessarily correspond to a live machine and some networks have firewalls that discard the UDP traceroute probes. Still, around 95% of the traceroute paths reach an address with the same origin AS as the target destination.

Unexpected final address: About 15% of the paths end in less than 30 hops (the default maximum number) with an address that differs from the intended destination. This can occur when the destination is a device (such as a router) that has multiple interfaces with different IP addresses, or if an intermediate component (such as a firewall) sends a PORT_UNREACHABLE message upon receiving unsolicited packets for a downstream host.

Ending with "*": More than half of the paths end with one or more "*" characters, implying that no ICMP reply was received. This can occur when the TTL-limited probes are discarded (say, by a firewall), the components along this part of the path do not participate in ICMP (or apply rate-limiting), or the ICMP messages are lost along the reverse path.

Ending with "!": Around 12% of the traceroute results end with a "!" symbol indicating that the last component in the path was unable or unwilling to forward the packets toward the destination. The two most common scenarios are !H (host unreachable) and !X (communication administratively prohibited), with !N (network unreachable) a distant third.

Ending after 30 hops with an IP address: About 7% of the paths continue to the maximum length (30 hops) and end with an IP address. The vast majority (95%) of these paths have forwarding loops, where some addresses appear multiple times in the path. A small fraction of the paths do not contain loops and appear to represent paths that continue beyond 30 hops.

Although most loops persisted till the end of the traceroute path, a few paths had temporary loops and some loops that ended with a "*". In total, 7–8% of the paths contained a forwarding loop. This may stem, in part, from IP addresses that have not been allocated to any operational network or machine. Some routers may be configured with default routes that direct the traffic back to an upstream router. We do not expect traceroutes to known "live" addresses to uncover such a large percentage of forwarding loops. Overall, the combination of forwarding loops, unreachable hosts, discarded probe packets, and devices with ICMP disabled resulted in a relatively small number of probes that traversed the entire path to the destination. To enable comparisons with the BGP data, we preprocessed the end of each traceroute path to remove forwarding loops and trailing "*" and "!" characters; then we converted the (partial) forwarding path to an AS path by mapping each hop to an AS number, where possible. As such, we did not expect a complete match between the BGP and traceroute AS paths. Instead, we compared the two paths up to and including the end of preprocessed traceroute path. For example, if the traceroute AS path is "4006 16631" and the BGP AS path is "4006 16631 22476," we considered this a successful match.

	Whois Data			Combined BGP Tables			Resolving Incompletes		
	AS 6431	AS 25	AS 3130	AS 6431	AS 25	AS 3130	AS 6431	AS 25	AS 3130
Match	44.7%	44.7%	46.1%	71.7%	73.20%	73.4%	77.8%	78.0%	81.6%
Mismatch	17.1%	29.4%	23.0%	6.1%	8.3%	7.2%	6.6%	9.0%	7.1%
Incomplete	38.2%	25.9%	30.9%	22.1%	18.5%	19.4%	15.6%	11.1%	11.3%
unmapped hop	33.4%	20.5%	25.9%	1.5%	2.7%	2.5%	0.3%	0.6%	0.3%
* hop	8.7%	7.2%	8.5%	9.1%	7.6%	8.7%	6.4%	4.6%	5.5%
MOAS hop	0.0%	0.0%	0.0%	13.0%	9.8%	9.7%	10.0%	6.9%	6.4%
Match/mismatch ratio	2.62	1.52	2.00	11.70	8.79	10.20	11.74	8.96	11.43

Table 5: BGP vs. traceroute AS paths for different AS mapping techniques

3.3 Comparing BGP and Traceroute Paths

To map IP addresses to AS numbers, we first applied the *whois.ra.net* data that form the basis of the "NANOG traceroute" tool [13]; the *whois.arin.net*, *whois.ripe.net*, and *whois.apnic.net* data were not appropriate for our purposes since these services do not provide the AS number associated with an IP address. Unfortunately, the *whois.ra.net* data are out-of-date and incomplete. The statistics in "Whois Data" columns in Table 5 show that the BGP and traceroute AS paths matched less than half of the time. Incorrect IP-to-AS mappings may be responsible for many of the "mismatches" with the BGP AS path. Many traceroute paths were "incomplete" because no mapping exists in the whois database for some of the router hops. Around 20–33% of the traceroute paths had "unmapped" IP addresses that *whois* could not associate with an AS; this is partially explained by ASes that have not updated *whois* to reflect their current address assignments.

To improve the IP-to-AS mapping, we combined BGP routing table data from many vantage points. Combining multiple routing tables provides (i) a richer view of different subnets that may be aggregated at other locations, (ii) a more complete picture of prefixes associated with multiple origin ASes, and (iii) a lower risk of missing certain prefixes due to transient reachability problems at any one router. Table 6 lists the number of prefixes in each BGP routing table, along with the number of prefixes with more than one origin AS. The RouteViews data [22] consisted of BGP routes learned from 23 participating ASes, mostly in the United States. The data from the RIPE-NCC Routing Information Service project [23] provided BGP routes from 75 ASes, mostly in Europe. The SingAREN routers [24] had BGP routes from ASes in the Asia-Pacific region. Each of the other tables provided BGP routes seen from one vantage point. All of the BGP tables were collected around May 29, 2003, in the middle of our traceroute experiments, to limit the effects of changes in the mapping of prefixes to origin ASes over time. Combining all of the tables produced a mapping with more than 200,000 prefixes and 16,000 ASes. About 10% of the prefixes mapped to multiple origin ASes.

Using the collection of BGP tables increased the "match" rate and substantially decreased the fraction of paths with "unmapped hops," as shown in the "Combined BGP Tables" columns in Table 5. This occurred because the BGP tables from the operational routers provide a more complete and up-to-date view of the "ownership" of the IP addresses appearing in the traceroute paths. Still, the BGP and traceroute AS paths agreed less than 73% of the time, even under our relatively liberal notion of "matching" (*i.e.,* after trimming the end of the traceroute paths). Less than 8.3% of the traceroute AS paths differ from the corresponding BGP AS path. In the remaining cases, the traceroute path was "incomplete" because one or more hops did not map directly to a single AS number:

Unmapped hop: In a few ($<$ 3.0%) of the paths, some hops had an address that did not match any prefix in the set of BGP tables. Private IP addresses accounted for less than 40% of the cases. Unmapped hops can arise when interfaces are assigned addresses that are not advertised to the larger Internet.

"*" hop: Many traceroute paths had one or more "*" characters, even after removing trailing "*" characters at the end of the path. A "*" hop may stem from a lost probe or ICMP packet, or from an intermediate node that does not participate in ICMP.

Multiple origin AS hop: Around 9–13% of the traceroute paths had at least one interface address that mapped to multiple AS numbers, making direct comparisons with the BGP path impossible. MOAS prefixes occur for various reasons including misconfiguration, multihoming, or exchange points [16].

The three cases are not mutually exclusive; a single traceroute path may have hops with one or more of these properties.

	Extracted Prefixes	Origin ASes	MOAS Prefixes
AS 6431	120997	15105	0
AS 25	124202	15213	0
AS 3130	121054	15086	0
AS 73	123583	15194	0
AS 6521	121096	15099	0
AS 14177	121135	15104	0
AS 10781	121669	15103	0
AS 22208	125050	15136	0
RouteViews	134095	15294	860
RIPE(00–08)	128960	15328	3400
SingAREN	6744	862	25
Potaroo	142348	16112	211
Verio	105381	13778	116
AT&T	128411	15171	109
Combined	203698	16367	8827

Table 6: BGP tables for IP-to-AS mapping around May 2003

4. RESOLVING INCOMPLETE PATHS

This section describes and evaluates three simple techniques that allow us to analyze a large fraction of the "incomplete" traceroute AS paths, as summarized in the "Resolving Incompletes" columns in Table 5. We discuss how to use internal router configuration files to validate the results, using data from a large service provider (AT&T, AS 7018) as an example. The internal configuration data enables us to verify whether certain interfaces belong to a particular AS and what lies on the other side of the link. We also can identify static routes that are used to direct traffic to specific customers. Our validation scripts could be used to compute statistics for other networks, without requiring these ASes to divulge their raw configuration files.

4.1 Unresolved Hops Within an AS

Many of the incomplete paths have "*" or unmapped hops in between two hops that map to the same AS; for example, a path may have one or more "*" hops between two interfaces that both map to AS 1239. We assume that such "*" and unmapped hops belong to the same AS as the surrounding hops; that is, we convert a path with hops "1239 * 1239" to a single AS-level hop of 1239. This is similar to the approach in [10] of clustering routers in a graph based on the AS number and associating each "*" interface with the nearest cluster. This simple heuristic reduced the number of incomplete paths with "*" hops by 30–40%. For unmapped hops, it reduced the incomplete paths by about 40%.

To test our hypothesis, we investigated the traceroute paths that appear to have one or more "*" hops within AS 7018 (*i.e.,* path segments such as "7018 * 7018" or "7018 * * 7018"). We inspect the IP address of the last hop in the path segment—the first hop after the "*" hops. We assume that this IP address corresponds to one end of the link from the previous router; the other end of the link should have the same network address. For example, a point-to-point link with the prefix 192.0.2.156/30 would have two interfaces with addresses of 192.0.2.157 and 192.0.2.158; 192.0.2.156 and 192.0.2.159 would correspond to the network and broadcast addresses, respectively. Upon seeing a hop with IP address 192.0.2.157, we look for another interface on a different router with IP address 192.0.2.158. In 98.1% of the cases, we are able to identify the router associated with this interface and verify that this router belongs to AS 7018. The remaining 1.9% of cases may have stemmed from transient routing changes where the hops in the traceroute path did not represent a single consistent path through the network.

4.2 Unmapped Hops Between ASes

Most of the unmapped hops appeared between interfaces that are mapped to different ASes (*e.g.,* "1239 ? ? 64"). We attempted to associate the unmapped hop(s) with the previous or subsequent AS, using DNS and *whois* data. First, we considered the suffix of the domain names associated with the interfaces (*e.g.,* converting "sl-gw9-ana-4-0-0.sprintlink.net" to "sprintlink.net"), including the country domain if present; reverse DNS lookups were successful for 59% of the IP addresses in the traceroute results. If an unmapped hop had the same DNS suffix as a neighboring (mapped) interface, we associated the unmapped hop with that AS. This is similar to the approach in [4] of using DNS names to identify routers belonging to the same service provider. However, DNS did not always return a name for the unmapped hop; if some other interface in the same /24 address block had a successful reverse-DNS lookup, we used the DNS suffix for that interface. Second, we used *whois* to identify the AS responsible for the unmapped interface; we used this AS mapping only when it matched one of the adjacent ASes in the traceroute path. These techniques reduced the number of paths with unmapped hops by over 50%; these "resolved" traceroute AS paths had about the same proportion of "matches" with the BGP AS paths as the initial "complete" traceroute paths did, increasing our confidence in these additions to the IP-to-AS mapping.

For this heuristic, validating with configuration data involved checking that each "?" hop mapped to AS 7018 actually corresponded to the IP address assigned to an interface in that network. However, the AS 7018 network numbers its interfaces out of an address block that is advertised to the rest of the Internet; as such, these interfaces did not appear as unmapped hops in the traceroute paths, and we could not use the configuration data to test the heuristic. Configuration data from other ASes would have been more useful. Overall, though, the fraction of "?" hops resolved by this set of heuristics was relatively low because we did not try to map "?" hops to other AS numbers (*e.g.,* besides adjacent AS hops like 1239 or 64). We experimented with other heuristics but did not believe that the DNS and *whois* IP-to-AS mappings were accurate enough to warrant a more liberal approach.

4.3 MOAS Hops at the End of the Path

Interface IP addresses that map to multiple origin ASes appeared in 10–13% of the traceroute AS paths. About 3% of all traceroute AS paths *end* with hops that map to multiple ASes. This can occur due to edge networks that connect to multiple providers without using BGP (or using private AS numbers) or due to misconfigurations [16]. We envision that an AS traceroute tool should report that these hops map to multiple ASes for diagnostic purposes. For the rest of the paper, we include these traceroute paths in our comparison with the corresponding BGP AS paths. We consider these traceroute hops a "match" with the corresponding BGP hop if the AS in the BGP path matches any *one* of the ASes associated with the traceroute MOAS hops. These "resolved" traceroute AS paths had about the same proportion of "matches" with the BGP AS paths as the initial "complete" traceroute paths.

Using the configuration data, we investigated the traceroute AS paths where the last hop was mapped to AS 7018 and at least one other AS. In particular, we inspected the IP prefixes used to map these hops to multiple origin ASes to see if they actually corresponded to customers of AS 7018. In all of the cases involving AS 7018, the prefix was specified in a static route associated with one or more access links to a customer. That is, AS 7018 originated the route to this prefix on behalf of a customer and, as such, the prefix referred not to equipment inside the backbone but rather to addresses in the customer's network.

5. IMPROVED IP-TO-AS MAPPING

After applying the techniques in Section 4, about 6–9% of the traceroute AS paths do not match the corresponding BGP AS path and another 6–10% have hops that map to multiple origin ASes. We suspect that inaccuracies in the IP-to-AS mapping are responsible for many of these cases. After a brief discussion of the causes of mismatches, we propose and evaluate algorithms for detecting IXPs, sibling ASes, and networks that do not announce routes for their infrastructure. The coverage of some of our techniques is limited by the fact that our measurement data come from only eight vantage points mostly in the United States, all directly connected to large providers in North America. The techniques discussed here and in the previous section are very efficient. The algorithms require on the order of a few minutes to run on traceroute paths to about 200,000 addresses.

5.1 Patterns and Causes of Mismatched Paths

At least two-thirds of the differences between the BGP and traceroute AS paths fell into one of four simple patterns:

Extra AS hop: For about 30–40% of the mismatches, the traceroute AS path had one extra intermediate hop that does not appear in the corresponding BGP AS path, as shown in Figure 2(a).

Missing AS hop: About 20% of the mismatches came from traceroute AS paths that were missing one intermediate hop compared to the BGP AS path, as shown in Figure 2(b).

Two-hop AS loop: Around 10% of the traceroute AS paths had an AS-level loop with two AS hops, such as the "H G" segment in Figure 2(c).

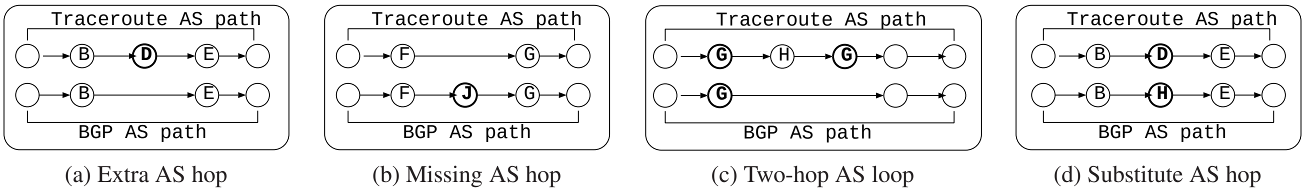


Figure 2: Mismatch patterns for the traceroute AS paths

	AS 6431	AS 25	AS 3130
Extra intermediate hop	33%	40%	41%
Missing intermediate hop	22%	20%	20%
Two-hop AS loop	9%	7%	8%
Substitute AS hop	3%	3%	2%
Other	33%	30%	29%

Table 7: Statistics on mismatched traceroute paths

	Extra	Miss	Loop	Subst	Other
Exchange point	X				
Sibling ASes	X	X	X	X	
Unannounced IP	X	X	X	X	
Aggregation/filtering					X
Inter-AS interface		X			X
ICMP source address	X	X		X	X
Routing anomaly	X	X	X	X	X

Table 8: Patterns and possible causes of mismatched AS paths

Substitute AS: In 2–3% of the cases, the two paths had a different AS for one intermediate hop, such as AS D for the traceroute path and AS H for the BGP path in Figure 2(d).

Table 7 summarizes the statistics, focusing on the first mismatch between each pair of AS paths. In each case, the "mismatch" between the two AS paths was nested within the path, starting with an initial matching hop.

Our heuristics look for common occurrences of these "differences" across many AS paths to identify possible mistakes in the IP-to-AS mapping applied to the traceroute AS paths. Finding multiple instances of each pattern increases the confidence in our explanation for why the paths differ and also makes our algorithms more robust to transient routing changes that may affect the accuracy of some of the traceroute paths. In practice, some traceroute paths may be affected by the results of multiple techniques, since we apply the improved IP-to-AS mapping across all of the traceroute paths. Our algorithms are based on the patterns we expect from common operational practices. Table 8 summarizes the seven root causes we consider, and the kinds of mismatch patterns they can create. The first three cases introduce mistakes in the IP-to-AS mapping and are the focus of this section. The remaining four cases are "legitimate" mismatches that do not necessarily stem from an incorrect mapping; we defer discussion of these cases to the next section. In practice, most of the items in Table 8 do not fall naturally into a single "mismatch pattern"; therefore, our algorithms need to look carefully across multiple instances of mismatch paths to draw meaningful conclusions.

5.2 Internet Exchange Points (IXPs)

IXPs are junction points where multiple service providers meet to exchange BGP routes and data traffic. An IXP typically consists of a shared infrastructure, such as an ATM switch or a FDDI ring, with physical connections to routers in each of the participating ASes. An IXP may have its own AS number and originate routes to its infrastructure; alternatively, the address of the shared infrastructure may be originated into BGP by one or more of the participating ASes. In either case, different pairs of service providers establish dedicated BGP sessions over the shared physical infrastructure. At the IP level, the forwarding paths traverse the shared equipment as shown in the left side of Figure 3. Yet, at the BGP session level, the participating service providers connect directly to each other, as shown in the right side of Figure 3. As a result, the AS-level forwarding path appears to have an extra AS hop relative to the corresponding BGP AS path, as shown earlier in Figure 2(a).

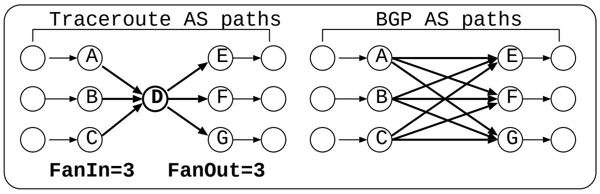


Figure 3: Traceroute vs. BGP AS paths through an IXP

The patterns in Figure 3 for the AS-level forwarding and signaling paths drive our algorithm for detecting IXPs. First, we inspect cases where the traceroute AS path has an extra hop compared to the corresponding BGP AS path; the extra hop could be a single AS D or an individual prefix that maps to multiple origin ASes. In practice, we do not expect to see the AS for an IXP to appear in any BGP AS paths, except as the origin AS for the paths for the shared equipment at the site. As such, the second step of our algorithm removes from consideration any AS D that appears as a transit AS in any BGP AS path. Finally, we expect an IXP to provide service to several pairs of ASes. As such, we check the number of unique ASes appearing just before and just after D; the example in Figure 3 has a fan-in and fan-out of 3. For robustness, we apply a threshold for the minimum fan-in and fan-out; in this paper, we apply a relatively small threshold of 2 since we only have measurement data from eight vantage points. Ideally, a larger threshold might be preferable for avoiding "false positives."

We also apply an additional requirement that for AS pairs consisting of the AS preceding and following the suspected IXP AS, there must at least two pairs with no AS in common. In other words, AS D is not considered as an IXP AS if it only appears as an extra AS in traceroute AS paths such as XDB and BDY, where X and Y are arbitrary ASes. As described in Section 5.3, AS B and D are likely to be siblings. This requirement is to assure the path diversity of selected IXPs and prevent mistaking a sibling AS for an IXP AS.

Applied to our measurement data, this algorithm found 477 cases (of an AS or a prefix) with a fan-in and fan-out of 1 or more with

	In	Out
California Research & Education Network (AS2151)	6	5
London IXP (AS5459)	4	7
Japan IXP (AS7527)	3	7
SANDY Network (AS5471)	2	2
PAIX (198.32.176.0/24)	9	50
Amsterdam IXP (193.148.15.0/24)	7	9
Seattle IXP (198.32.180.0/24)	6	32
Chicago Ameritech (206.220.243.0/24)	4	37
Equinix IBX San Jose (206.223.116.0/24)	4	20
Japan IXP (JPIX) (210.171.224.0/24)	4	9
London IXP (LINX) (195.66.224.0/19)	4	7
Hong Kong IXP (HKIX) (202.40.161.0/24)	4	6
Equinix Ashburn (206.223.115.0/24)	3	7
Tokyo Network Service Provider IXP (202.249.2.0/24)	3	5
Western Australia (WAIX) (198.32.212.0/24)	3	2
Hutchison Telecommunications, HK (210.0.251.0/24)	3	2
MAE West ATM San Jose (198.32.200.0/24)	2	13
Equinix IBX Secaucus (206.223.117.0/25)	2	4
MAE East (198.32.187.0/24)	2	3
Japan Network Information Center (202.249.0.0/17)	2	3
SI-TELEKOM-193-77, Slovenia (193.77.0.0/16)	2	3
Mae-West Moffet Field (198.32.136.0/24)	2	2
Lipex Ltd, Telehouse Network, UK (193.109.219.0/24)	2	2
Comite Gestor da Internet no Brasil (200.187.128.0/19)	2	2
ROSTELECOM-NET, Russia (213.24.0.0/16)	2	2

Table 9: AS numbers and prefixes inferred as IXPs

corresponding AS appearing in traceroute AS paths but not BGP paths. Only 25 cases had fan-in and fan-out of at least 2 and satisfy our criteria of an IXP; these cases are listed in Table 9 in decreasing order of fan-in and fan-out. To verify our results, we first queried *whois* using the AS number or prefix to see if the description contained the words "exchange point" or "Internet exchange"; for example, AS 5459 was listed as "London Internet Exchange" in *whois.ripe.net*. This check succeeded for 18 of our 25 inferences. Then, we compared our results against a list of known IXPs [25]. This confirmed 16 of the 25 inferences. Together, 19 of the 25 inferences passed at least one of these checks. Some of the remaining cases (highlighted in italics) may be IXPs, too; for example, CalRen is an exchange point for universities in California.

Inspecting the list of known IXPs, we find that we missed 13 known IXPs. Among them, all but one had a *fan-in* of 1; for example, the PAIX Seattle exchange point had a fan-in of 1 and a fan-out of 5. The 13 cases include 2 NAPs (in Seattle and Miami), 4 European IXPs, 1 Asian IXP, 2 Equinix sites, and 4 small IXPs in the exchange point block 198.32.0.0/16. We believe that our algorithm missed these cases due to the small number of measurement locations; in addition, our measurement sites connect directly to large tier-1 providers in the U.S. except for one site connecting to a large provider in Canada, limiting the number of ways the traceroute paths could reach the IXPs. In the end, some of these remaining IXPs are potentially mistakenly placed in other categories by the techniques described later in this section.

Using the list of IXPs generated by our algorithm, an AS-level traceroute tool could indicate which IP-level hops map to exchange points. We used our results to map these IP addresses to null ASes; that is, we remove the IXP ASes and prefixes from the traceroute AS paths. For example, a traceroute AS path with "$B\ D\ E$" would become "$B\ E$" after removing AS D. The results of applying the new IP-to-AS mapping across all of the traceroute paths is shown in the "Internet Exchange Points" columns in Table 10. Compared with the earlier results in Table 5, the number of matched paths increased to 78.2-85.4%, corresponding to an increase of 1–4 percentage points. This occurs due to a decrease in both the number of mismatched paths and the number of incomplete paths. For the AS 6431 data, the IXP algorithm resolved more than half of the incomplete paths with MOAS hops. We would expect more dramatic results for sites that connect to smaller providers that tend to route more of their traffic through IXPs rather than private peering links.

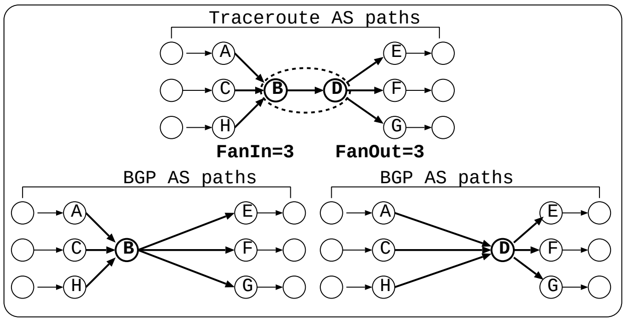


Figure 4: Traceroute and BGP AS paths with siblings

5.3 Sibling ASes

In some cases, a single organization owns and manages multiple ASes, sometimes as a result of mergers and acquisitions. The ASes may share address space, with one AS numbering some of its equipment using part of an address block originated by another. This affects the mapping of traceroute hops to AS numbers, and can lead to ambiguity about which AS actually carries the traffic; in some sense, the distinction between the two ASes may not be important since they "belong together." In the example at the top of Figure 4, the traceroute AS paths includes ASes B and D though the BGP AS path includes only one of the two ASes, as shown in the bottom of the figure. This phenomenon can result in traceroute AS paths that have an extra AS hop (B or D) relative to the corresponding BGP paths. Sibling ASes can also produce traceroute paths with other patterns, as discussed in the next subsection.

The patterns in Figure 4 suggest a way to identify cases where sibling ASes affect the traceroute AS path. Similar to the IXP algorithm, we consider the fan-in and fan-out of traceroute AS paths traversing a two-hop segment "$B\ D$" that corresponds to a single AS hop in the corresponding BGP paths. For robustness, we apply a threshold to the fan-in and fan-out; in this paper, we enforce a minimum fan-in and fan-out of two. In addition, we focus on cases where one of the two ASes (say, AS D) never appears in a BGP AS path, except as an origin AS. That is, we assume that one AS (B) is using the address space originated by the other AS (D), rather than trying to capture cases where each AS borrows from the other.

In applying this algorithm to our data, we identified 28 pairs of sibling ASes. The fan-in and fan-out were as large as 10 and 31, respectively. To check our results, we inspected the *whois* entries for the ASes and found that in 15 cases the two ASes had the same organization name (*e.g.,* ASes 1239 and 1791 belonged to Sprint and ASes 1299 and 8233 belonged to TeliaNet). In the remaining seven cases, the AS pairs appeared together as originating ASes for one or more prefixes in the BGP routing tables, adding extra credibility to the conclusion that they are siblings. As part our future work, we plan to compare our sibling inferences with the results of algorithms for inferring AS relationships from BGP AS paths [19, 20].

We modified the IP-to-AS mapping based on these results to treat

	Internet Exchange Points			Sibling ASes			Unannounced Addresses		
	AS 6431	**AS 25**	**AS 3130**	**AS 6431**	**AS 25**	**AS 3130**	**AS 6431**	**AS 25**	**AS 3130**
Match	78.2%	84.4%	85.4%	86.0%	85.9%	87.0%	90.0%	90.6%	91.0%
Mismatch	6.4%	8.7%	7.1%	6.4%	7.8%	6.2%	2.7%	3.5%	2.6%
Incomplete	15.4%	6.9%	7.5%	7.6%	6.3%	6.8%	7.4%	6.0%	6.6%
Match/Mismatch ratio	12.20	9.70	12.06	13.42	11.00	14.08	33.51	25.95	35.41

Table 10: The results of using the three techniques to tune the IP-to-AS mapping

sibling ASes as a single network. That is, we replaced every occurrence of B or D in the IP-to-AS mapping with the set $\{B, D\}$. We considered the traceroute and BGP AS hops a "match" if the BGP AS hop was the same as either of the two siblings in the traceroute AS path. After applying the new IP-to-AS mapping to all of the traceroute paths, 85.9-87.0% of the traceroute AS paths matched the corresponding BGP AS paths. This increase came from up to a 12% reduction in the mismatched paths and up to a 50% reduction in the incomplete paths. As a result, the mismatched and incomplete paths became as low as 6.2% and 6.3% of the total number of paths, respectively, as shown in the "Sibling ASes" columns of Table 10.

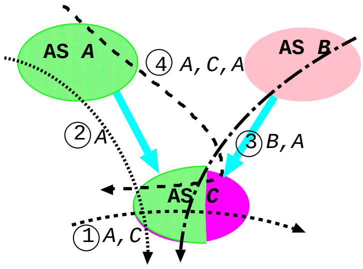


Figure 5: Mismatches caused by unannounced IP addresses

5.4 Unannounced Infrastructure Addresses

An AS does not necessarily announce the addresses assigned to its equipment via BGP. This can lead to "unmapped" addresses, as discussed earlier in Section 3.3. However, sometimes these addresses fall into larger address blocks originated by the AS's sibling or provider. This can cause several patterns of mismatches between the BGP and traceroute AS paths. In the example in Figure 5, AS C connects to two upstream providers A and B. AS A has allocated a subnet of its address space to AS C and originates the supernet in BGP to the rest of the Internet. AS C uses its part of the address block to number some of its equipment but C does not advertise the subnet in BGP. As a result, some traceroute hops in AS C are mistakenly mapped to AS A. Figure 5 shows four example paths:

Extra hop: Path 1 traverses some hops in AS C that (mistakenly) map to A and others that (correctly) map to C, resulting in a traceroute path of "A C" rather than "C".

Missing hop: Path 2 traverses both A and C, resulting in a BGP path of "A C." However, the hops in C are (mistakenly) mapped to A, resulting in a traceroute path of "A".

Substitute hop: Path 3 traverses both B and C, resulting in a BGP path of "B C." However, the hops in C are (mistakenly) mapped to A, resulting in a traceroute path of "B A."

AS loop: Path 4 traverses ASes A and C, resulting in a BGP path of "A C." However, some of the hops in C are (mistakenly) mapped to A, resulting in a traceroute path of "A C A."

Focusing first on AS loops, our algorithm looks for the loop patterns in Figure 6(a). We count the number of times ASes G and H appear together in this pattern, where the traceroute AS path has a loop and the corresponding BGP path has a single hop for each AS. In analyzing our data, we found that small number of AS pairs appeared in many such paths, and these accounted for the vast majority of the loops. Our algorithm applies a threshold of 50 occurrences before inferring that ASes G and H "share" address space and changes the mapping of the second G hop to an H; that is, once a traceroute AS path appears to "enter" an AS H, we assume that the path continues in this AS. In effect, we assume that H "owns" the addresses of these traceroute hops but did not advertise them in BGP. However, we do not know the size of the address block allocated to H. We inspect the IP addresses of the individual traceroute hops involved and add the corresponding /24 prefix to our IP-to-AS mapping (with H as the associated AS). In applying this method, we found 20 unique AS pairs responsible for 830 unannounced /24 prefixes; many of these prefixes were adjacent, suggesting that some larger subnets were involved. Furthermore, the matched prefixes of the corresponding IP addresses tend to have shorter length, indicating that there may be smaller subnets missing in our prefix to AS mapping.

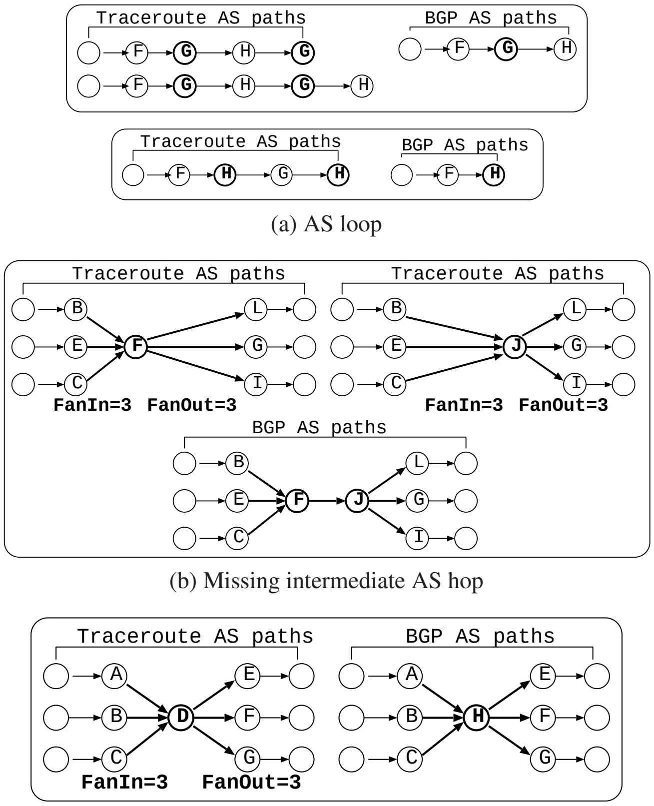
To check our results, we inspected the *whois* entries for these ASes and confirmed that in half of the 20 cases the two ASes belonged to the same institution (*i.e.,* the two ASes are siblings). In two other cases, the AS pairs could be classified as siblings based on their Web sites—AS 174 (PSINet) and AS 16631 (Cogent Communications), and AS 209 (Qwest) and AS 3908 (Supernet). These two examples are cases where the *whois* data do not capture acquisitions or mergers. Six more cases appeared to have a provider-customer relationship, in that *whois* showed one AS (the "customer") responsible for a subnet of an address block assigned to the other AS (the "provider"). In these cases, *whois* had address assignment information that was not available from the BGP routing tables since the "customer" subnet was not visible in any of our datasets. We were unable to verify the remaining two AS pairs.

For extra and substitute ASes, we follow a similar approach to the algorithms for IXPs and siblings. Focusing on patterns like Figure 3 and Figure 6(c), we apply a threshold of fan-in and fan-out of two to infer that an AS pair "shares" address space. Unlike the IXP and sibling algorithms, we apply these checks at the *prefix* level, assuming that some /24 prefix that has not been announced. For the "extra hop" case, we identified 308 such /24 prefixes; for the "substitute hop" case, we identified 25 prefixes. The case of a "missing hop," shown in Figure 6(b), is more complicated. By applying the fan-in and fan-out thresholds, we identified 77 AS pairs that appeared to "share" address space. However, we do not have a reliable way to determine which parts of the address block should be associated with the "missing" AS. Therefore, we do not use these results to modify our IP-to-AS mapping in any way. In ongoing work we are exploring ways to handle "missing" hops.

After identifying the unannounced addresses and the owning AS,

	AS 6431		AS 25		AS 3130	
Number of vantage points	3	8	3	8	3	8
Match	88.5%	90.0%	89.2%	90.6%	88.5%	91.0%
Mismatch	4.0%	2.7%	4.7%	3.5%	3.8%	2.6%
Incomplete	7.5%	7.4%	6.1%	6.0%	6.7%	6.6%
Match/Mismatch ratio	22.11	33.51	18.89	25.95	22.99	35.41

Table 11: The effect of multiple vantage points: comparing using the first three with all eight probing locations.



(a) AS loop

(b) Missing intermediate AS hop

(c) Substitute intermediate AS hop

Figure 6: ASes not announcing their infrastructure addresses

we modify the IP-to-AS mapping to add a new entry for each /24 prefix. Applying the new IP-to-AS mapping across all of the traceroute paths reduced the number of mismatched paths by as much as a factor of two. In addition, the new mapping slightly reduced the fraction of incomplete paths. Ultimately, after applying all three of the techniques in this Section, the "match" rate exceeded 90% for each data set and the ratio of matches to mismatches ranged from 25-35. Still, a small fraction (2.6–3.5%) of the traceroute AS paths did not agree with the BGP AS paths; Section 6 explores possible explanations for the remaining mismatches.

5.5 Diversity of Probing Locations

Our techniques rely on the topology diversity of the traceroute measurements. Increasing probing locations increases the likelihood that a different AS-level path is used to traverse pairs of siblings, Internet eXchange Pointis, and unannounced address spaces. This, in turn, reduces the probability that they would be missed in an AS-level traceroute tool based on our techniques. Both the geographic location and the upstream connectivity have an impact on the diversity of AS-level paths. Previous work [26] studied the marginal utility of discovering network topology using traceroute. They concluded that increasing the number of sources in traceroute experiments has low utility beyond the second source. Increasing the number of sources is admittedly more important for our purposes, though, since our heuristics rely on fan-in as well as fan-out counts.

In our study, we try to cover all the destination prefixes in the local BGP table. For each source, the set of destination probed is roughly the same. We found that adding additional sources in our study significantly increases the fan-in and fan-out counts across both sibling and IXP ASes. We compare the inference results based on measurements from the first three vantage points with all eight locations. For example, the fan-in and fan-out count going through PAIX, the Palo Alto Internet eXchange Point, increased from 5 and 14 to 9 and 50 respectively. Four known IXPs (Equinix San Jose, London IXP, Mae-West San Jose, and Mae-East) were missed using the first three locations due to insufficient fan-in and fan-out count, but they are correctly inferred using all eight data sets. As several newly added locations are in California, exchange points in San Jose are therefore more likely to be inferred.

Table 11 compares the match between traceroute AS paths and BGP AS paths using data from the first three locations with the complete data from all eight locations. The improvement is due to newly discovered IXPs, siblings, and unannounced address blocks as result of increased path diversity. The increase in matched paths is only between 1.5 and 2.8%; however, the reduction in mismatched paths ranges between 25–30%. This eliminates the false positives for potential routing problems that network operators need to investigate further. The table also shows that the match to mismatch ratio of comparing local BGP table AS paths with traceroute AS paths increased by 35–50%. We believe that adding vantage points in Europe and Asia would offer further advantages.

6. LEGITIMATE AS PATH MISMATCHES

In this section, we discuss four "legitimate" reasons why the traceroute and BGP AS paths may disagree, and speculate on whether the cases might explain some of the remaining "mismatches." Where possible, we look for evidence of these cases in our routing data and in the configuration files for AS 7018. We also propose additional measurement that would help classify these mismatches more precisely.

6.1 Route Aggregation /Filtering

At each of our eight measurement locations, the local BGP table does not have a complete view of the IP prefixes throughout the Internet. To limit protocol and storage overhead, routers may be configured to filter routes for certain subnets or combine multiple subnets together into a single aggregated route [27]. For example, Figure 7 shows an AS C that has the address block 8.0.0.0/8 and

	AS 6431	AS 25	AS 3130
Extended path	22%	18%	19%
Missing hop	24%	25%	27%
Extra hop	9%	12%	13%
Other	45%	45%	41%

Table 12: Remaining mismatches with BGP AS path

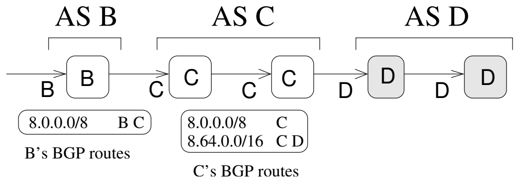


Figure 7: Extended traceroute path due to filtering by AS C

assigns the subnet 8.64.0.0/16 to its customer, AS D. Although AS C has BGP routes for both prefixes, only the route for 8.0.0.0/8 is propagated to AS B. Packets from AS B to the destination 8.64.0.1 would have a longest-matching prefix of 8.0.0.0/8 (with an AS path of "B C" in the local BGP routing table). However, the forwarding path would actually continue beyond AS C through one or more hops in AS D. Whether these traceroute hops are mapped correctly to AS D depends on whether the addresses of D's interfaces (which may or may not fall within the 8.64.0.0/16 block) are announced into BGP and are seen from the vantage points where we collect BGP routing tables.

Since many of our traceroute experiments do not traverse the entire forwarding path to the destination, we may significantly *undercount* the cases where route aggregation results in a BGP AS path that "ends early" relative to the forwarding path for destinations in a smaller (unseen) subnet. Yet, across the three data sets, extended traceroute AS paths still account for 18–22% of the mismatches with the BGP AS paths, as shown in Table 12. To test our hypothesis that route aggregation is responsible for some of these cases, we compare the AS-level forwarding paths for the two IP addresses in each prefix (*e.g.,* 8.0.0.1 and 8.64.0.1). Across all of the prefixes where both forwarding paths are "complete," the two IP addresses have the same AS-level forwarding paths more than 99% of the time. However, when we focus on cases when either (or both) of these IP addresses has an "extended" path, this number drops below 75%; in more than 20% of the cases, one address has a forwarding AS path that matches the BGP AS path and the other has an extended path. The differences in the pairs AS-level forwarding paths are consistent with the effects of route aggregation/filtering.

6.2 Interface Numbering at AS Boundaries

Traceroute reports the IP addresses of *interfaces* rather than routers. In practice, interfaces to the same link are assigned addresses from the same prefix (*e.g.,* interfaces 192.0.2.157 and 192.0.2.158 forming a single point-to-point link with prefix 192.0.2.156/30). This introduces a potential problem for a link between two ASes—the interfaces are typically assigned an address block belonging to one of the two ASes, not both. In some cases, the path may enter and leave a router in some AS C where the two hops have addresses "owned" by the adjacent ASes, such as B and D, as shown in Figure 8. In this example, the traceroute AS path appears to have a segment "B D" when the path actually traverses a single router in AS C; in contrast, the "B C D" in the BGP AS

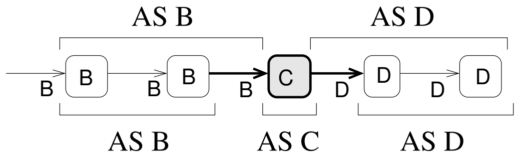


Figure 8: Missing AS hop C due to interface numbering

path is correct. As such, interface numbering at AS boundaries can result in a traceroute AS path that has a "missing" AS hop when compared to the corresponding BGP AS path. About 25–27% of the remaining "mismatches" between the BGP and traceroute AS paths stem from a single "missing" AS hop in the traceroute path, as shown in Table 12; we speculate that interface numbering at AS boundaries may be partially responsible.

To quantify these effects, we inspected cases where AS 7018 appeared as AS B, C, or D in a BGP path where the corresponding segment of the traceroute path "B D." AS 7018 never appeared as AS D and appeared only once as AS C; as such, we focused our attention on the case where AS 7018 corresponded to AS B. We first extracted the IP address of the last hop in the traceroute path that mapped to AS 7018; then, we generated the IP address of the other end of the link (*e.g.,* converting 192.0.2.158 to 192.0.2.157) and looked for an interface with this IP address in the configuration files from the same day. Then, we looked in the same configuration file to see if the interface was associated with a BGP session to a neighboring domain; if so, we extracted the remote AS number associated with this BGP session and compared it to the AS C in the BGP AS path. In more than 97% of the cases, we found that the last hop in AS 7018 was an interface associated with a BGP session to AS C rather than AS D or any other AS. In Section 7, we discuss how router-level graphs of the Internet [4, 5, 6] could help resolve these kinds of ambiguities.

6.3 Outgoing Interface in ICMP Message

Traceroute "discovers" a hop along the forwarding path from the source address of the ICMP TIME_EXCEEDED message sent in response to a TTL-limited probe. Ideally, the address corresponds to the *incoming* interface where the packet entered the router. However, the ICMP RFC [28] does not explicitly state which IP address the router should use. In practice, some routers may assign the source address based on the *outgoing* interface used to forward the ICMP message back to the host initiating the traceroute [11]. Since routing is not necessarily symmetric, the interface receiving the traceroute probe and the interface sending the ICMP message are not always the same. When this happens, traceroute reports the wrong forwarding path which can, at times, result in an incorrect AS-level path. Figure 9 shows an example where the actual forwarding path traverses ASes B and D, though traceroute reports an incorrect hop that maps to AS C. This can result in a traceroute AS path with "B C D" when the corresponding BGP path is simply "B D." About 9–13% of the remaining "mismatches" have a single extra AS hop in the traceroute AS path; we speculate that ambiguity about the source IP address in the ICMP reply may be responsible for some of these cases.

The work in [11] checked the source code for several IP stacks and tested the behavior of a Cisco 7500 router; only the Linux IP stack used the address of the outgoing interface in the TIME_EXCEEDED message. We evaluated several other popular commercial routers and operating system versions in our test lab. Routers using the address of the incoming interface included the Cisco GSR (IOS 12.0(21)S3), Cisco 7200 (IOS 12.2.(10a)), Juniper

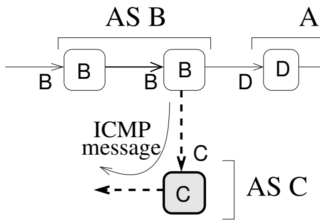


Figure 9: Extra AS hop C due to outgoing interface in ICMP

M10 (JunOS 5.3R2.4), and Avici TSR (4.2.1A); however, the Cisco 3660 running IOS 12.0(7)XK1 used the IP address of the outgoing interface in its TIME_EXCEEDED replies. From our tests and the results in [11], we believe that the outgoing-interface problem might affect some of the traceroute paths, particularly for hops in smaller ASes that use lower-end routers. Determining whether this phenomenon explains some of our "extra intermediate hop" cases is difficult in practice. Ultimately, additional active measurements may be necessary to probe a suspicious router from multiple vantage points to infer its behavior.

6.4 Routing Anomalies

When the underlying route is changing, the "hops" returned by traceroute do not necessarily represent a single path through the network. This problem arises because each hop in the traceroute output corresponds to a separate TTL-limited probe that might not traverse the same forwarding path as the other probes sent toward the destination. In our preprocessing, we eliminated traceroute experiments where the corresponding BGP-level path was changing, so we may not see as many cases where routing changes occur. Still, the forwarding path may fluctuate even if the BGP path does not. Intradomain routing would tend not to alter the AS-level path but we cannot dismiss this possibility entirely. In addition, the AS-level forwarding path may change if some downstream AS experiences a BGP routing change for some subnet of the advertised prefix. To increase our confidence in the forwarding path, we could repeat the traceroute experiments in cases where the BGP and traceroute AS paths disagree to make sure that transient changes in the forwarding path are not to blame.

In addition, some routing anomalies can cause the forwarding and signaling paths to differ even when both are stable. This can arise due to "deflections," where a router directs a packet to an intermediate node that has a different view of the "best" BGP route for a destination. The work in [12] describes how certain internal BGP (iBGP) configurations can be vulnerable to deflections; these scenarios would be extremely difficult for an operator to detect and debug. In many cases, a deflection would not change the AS-level path since the "best" AS path at different points in the network might exit via the same neighboring AS. Still, in some cases the two routers may pick different (equally good) best paths, such as AS B selecting a path through AS C (*e.g.,* "C E F") at one peering point and a path through AS D (*e.g.,* "D G F") at another. In such situations, deflections may cause the packets to traverse one of these paths despite the router having a BGP table with the other route. These kinds of anomalies could produce a variety of patterns in how the BGP and traceroute AS path differ.

7. CONCLUSIONS

In this paper, we have proposed techniques for improving how IP addresses of network infrastructure are mapped to the administering ASes. These techniques rely on a measurement methodology for (i) collecting both BGP and traceroute paths at multiple vantage points and (ii) using an initial IP-to-AS mapping derived from a large collection of BGP routing tables. We proposed simple heuristics for resolving traceroute paths with "*" and unmapped IP-level hops and describe how to verify the results using internal configuration data. Then, we presented heuristics that compare the BGP and traceroute AS paths to identify IXPs, sibling ASes, and other ASes that "share" address space, and evaluated the improved IP-to-AS mapping on traceroute paths collected from three vantage points. Compared to an initial IP-to-AS mapping constructed from the BGP tables, our heuristics reduced the fraction of incomplete paths from 18–22% to 6–8%; the ratio of matched to mismatched paths more than doubled, increasing from around 9–12 to 25–35. The adjustments to the IP-to-AS mapping are crucial for building an accurate AS-level traceroute tool for network operators and researchers. In addition, the improved mapping helps in highlighting the small number of important cases when the traceroute and BGP AS paths actually differ.

Our techniques capitalize on certain operational realities which arguably could change over time. For example, we were able to include more than 99% of the BGP AS paths in our analysis because most BGP routes are relatively stable and few BGP AS paths have private ASes or AS_SETs. We also exploited the fact that most ASes assign public, routable addresses to their equipment and often give meaningful domain names to the interfaces. Although quite a few traceroute hops did not return ICMP replies, most of the "*" hops occurred near the ends of paths or between other hops in the same AS. In addition, our techniques build on the assumption that the AS-level signaling and forwarding paths typically (though not always) match. This assumption would become less reasonable if route filtering were applied more aggressively in the core of the Internet, or if routing anomalies such as deflections were very common. Also, if the practice of "multi-homing without BGP" becomes more common, the notion of "origin AS" would become increasingly ambiguous. We plan to investigate the sensitivity of our results to these factors.

Converting an IP-level path to an AS-level path is extremely difficult, and additional measurement data would help. An accurate router-level graph [18, 4, 5, 6] would allow us to map interfaces to routers and, in turn, map routers to ASes. This would make our techniques less vulnerable to the interface numbering at AS boundaries (Section 6.2) and the source IP address in ICMP messages (Section 6.3). Although challenging in its own right, collecting the router-level topology does not require joint collection of BGP update messages, expanding the set of possible locations for launching the necessary traceroute probes. Our efforts would benefit from collecting both traceroute and BGP data at more locations, particularly in Europe and Asia. We are working on expanding the number and diversity of locations where we collect our data. Also, we are exploring the use of the public traceroute servers despite the many challenges they introduce. In particular, we are investigating ways to reduce the amount of measurement data needed from each vantage point to lower the load we would impose on the public servers.

Ultimately, developing an accurate AS traceroute tool depends on having a platform for collecting and managing information about the Internet infrastructure. Having a generic distributed platform, supported by service providers, for collecting and combining the traceroute and BGP data would be extremely valuable. Going one step further, computing the AS-level traceroute path would be much easier if ASes kept an up-to-date list of the address blocks used to number their equipment. This would simplify the interpretation of the source addresses in the ICMP messages. ASes could still protect access to their infrastructure from possible attack by filtering packets and routes that refer directly to their equipment. Alternatively, the ICMP specification could be extended to include an AS number or other identifying information in ICMP replies. In addition, the ICMP specification could be augmented to clarify whether the source address of the ICMP response messages refers to the incoming or outgoing interface at the router.

In our ongoing research, we are working on a public-domain AS traceroute tool that exploits our improved IP-to-AS mapping. We plan to use the tool to develop techniques for real-time detection and diagnosis of routing anomalies.

Acknowledgments

We would like to thank Jay Borkenhagen, Tim Griffin, Michael Rabinovich, Shubho Sen, Aman Shaikh, and Hoi-Sheung Wilson So for their valuable comments on the paper. We would also like to thank Randy Bush for answering questions about operational practices and Michael Rabinovich for a useful suggestion on the sibling heuristic. Thanks also to Joel Gottlieb for his help in working with the router configuration data. This work would not have been possible without the access to local BGP data and machines for performing traceroute made available by Dave Andersen, Randy Bush, Nick Feamster, Tim Griffin, John Hess, Ratul Mahajan, the MIT RON project and the PlanetLab project. Thanks also to RouteViews and RIPE-NCC for making their BGP routing tables and update messages available to the research community. Finally, we thank our shepherd David Tennenhouse and anonymous reviewers for their help to improve the paper.

8. REFERENCES

[1] Van Jacobson, "Traceroute," `ftp://ftp.ee.lbl.gov/traceroute.tar.gz`.

[2] V. Paxson, "End-to-End Routing Behavior in the Internet," *IEEE/ACM Trans. Networking*, vol. 5, no. 5, pp. 601–615, October 1997.

[3] Stefan Savage, Andy Collins, Eric Hoffman, John Snell, and Tom Anderson, "The end-to-end effects of Internet path selection," in *Proc. ACM SIGCOMM*, September 1999.

[4] Ramesh Govindan and Hongsuda Tangmunaraunkit, "Heuristics for Internet map discovery," in *Proc. IEEE INFOCOM*, 2000.

[5] "Skitter," `http://www.caida.org/tools/measurement/skitter`.

[6] Neil Spring, Ratul Mahajan, and David Wetherall, "Measuring ISP topologies with Rocketfuel," in *Proc. ACM SIGCOMM*, August 2002.

[7] Paul Barford, Azer Bestavros, John Byers, and Mark Crovella, "On the marginal utility of network topology measurements," in *Proc. Internet Measurement Workshop*, November 2001.

[8] Ratul Mahajan, David Wetherall, and Tom Anderson, "Understanding BGP misconfigurations," in *Proc. ACM SIGCOMM*, August 2002.

[9] "Visualizing Internet topology at a macroscopic scale," `http://www.caida.org/analysis/topology/as_core_network/`.

[10] Hongsuda Tangmunarunkit, Ramesh Govindan, Scott Shenker, and Deborah Estrin, "The impact of policy on Internet paths," in *Proc. IEEE INFOCOM*, 2001.

[11] Lisa Amini, Anees Shaikh, and Henning Schulzrinne, "Issues with inferring Internet topological attributes," in *Proceedings of SPIE*, July 2002, vol. 4865.

[12] Timothy G. Griffin and Gordon Wilfong, "On the correctness of iBGP configuration," in *Proc. ACM SIGCOMM*, August 2002.

[13] "Nanog traceroute," `ftp://ftp.login.com/pub/software/traceroute/`.

[14] "Prtraceroute," `http://www.isi.edu/ra/RAToolSet/prtraceroute.html`.

[15] Paul Barford and Winfred Byrd, "Interdomain routing dynamics," Unpublished report, June 2001.

[16] Xiaoliang Zhao, Dan Pei, Lan Wang, Dan Massey, Allison Mankin, S. Felix Wu, and Lixia Zhang, "An analysis of BGP multiple origin AS (MOAS) conflicts," in *Proc. Internet Measurement Workshop*, November 2001.

[17] Young Hyum, Andre Broido, and k claffy, "Traceroute and BGP AS Path incongruities," 2003. `http://www.caida.org/outreach/papers/2003/ASP/`.

[18] Hyunseok Chang, Sugih Jamin, and Walter Willinger, "Inferring AS-level internet topology from router-level path traces," in *Proc. Workshop on Scalability and Traffic Control in IP Networks, SPIE ITCOM Conference*, August 2001.

[19] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. Networking*, December 2001.

[20] Lakshminarayanan Subramanian, Sharad Agarwal, Jennifer Rexford, and Randy H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *Proc. IEEE INFOCOM*, June 2002.

[21] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed Internet routing convergence," *IEEE/ACM Trans. Networking*, vol. 9, no. 3, pp. 293–306, June 2001.

[22] "University of Oregon Route Views Project," `http://www.routeviews.org/`.

[23] "Ripe NCC," `http://www.ripe.net/ripencc/pub-services/np/ris/`.

[24] "Singaren," `http://noc.singaren.net.sg/netstats/routes/`.

[25] "Packet Clearing House," `http://www.pch.net/resources/data/exchange-points/`.

[26] Paul Barford, Azer Bestavros, John Byers, and Mark Crovella, "On the Marginal Utility of Network Topology Measurements," in *Proc. Internet Measurement Workshop*, November 2001.

[27] E. Chen and J. Stewart, "A Framework for Inter-Domain Route Aggregation," Request for Comments 2519, February 1999.

[28] J. Postel, "Internet Control Message Protocol," RFC 792, September 1981.