Traceroute and BGP AS Path Incongruities

Young Hyun, Andre Broido, kc claffy
CAIDA, San Diego Supercomputer Center,
University of California, San Diego
E-mail: {youngh, broido, kc}@caida.org *

Abstract

Researchers investigating topics such as performance, stability, and growth of the Internet often turn to BGP routing tables to obtain Internet topology. BGP routing tables provide a mapping from address prefixes to autonomous system (AS) paths. Our study, based on hundreds of thousands of traceroutes from three locations worldwide, categorizes differences between AS paths obtained from BGP routing tables and AS paths derived from traceroute paths. We find much of the disparity results from *exchange points ASes* (which rarely appear in BGP paths) and by groups of *ASes under the same ownership*. We introduce a new AS relationship, *common ownership*, that reflects the complexities of real-world business relationships and practices. We conjecture that the observed difference in size between the cores of an AS graph derived from BGP and an AS graph derived from traceroute is due to the visibility of peering at exchange points in traceroute paths.

1 Introduction

Connections between participants in Internet communications can be abstracted in two possible ways. One is a five-layered model of the protocol stack; another is in the dimension of network administration. The second abstraction groups IP addresses into subnets, subnets into network prefixes and prefixes into autonomous systems (ASes). An autonomous system is the term in Border Gateway Protocol (BGP) [1] for an entity that manages one or more networks and has a coherent policy for routing IP traffic both internally and to other ASes.

Each aggregation step in the sequence outlined above involves a decimal order of magnitude reduction in complexity of its description, measured by the number of objects of each kind. Thus 10^6 IP addresses become 10^5 prefixes and 10^4 ASes. Most analysis of Internet connectivity is done at the AS level.

In this paper we investigate how closely the concept of an AS matches a faithful, yet concise, representation of Internet topology. We compare two types of data that predominate in research on AS-level Internet connectivity: BGP AS paths and traceroute AS paths. Accurate knowledge and representation of Internet topology is essential for understanding why most BGP route changes converge to a stable state within a reasonably short amount of time (minutes), although theoretical analysis [2] and empirical observation [3] show that timely convergence need not always occur.

Previous work on AS connectivity has been aimed at constructing AS graphs, that is, connections from an AS to its immediate neighbors. This structural framework facilitates analysis of what is possible in Internet traffic routing, yet is too coarse to measure what really occurs since it does not account for policy constraints or the variety of roles played by ASes with respect to each other.

In [4], we introduced the notion of the combinatorial core of a graph and showed that the core of skitter AS graphs is much larger than the core of AS graphs obtained from RouteViews tables even though the number

*Support for this work is provided by the Defense Advanced Research Project Agency (DARPA) NMS (N66001-01-1-8909) program and DISA's National Communiations System Organization. CAIDA is a collaborative organization supporting cooperative efforts among the commercial, government and research communities aimed at promoting a scalable, robust Internet infrastructure. CAIDA is based at the University of California's San Diego Supercomputer Center (SDSC). www.caida.org.

of AS nodes and links is comparable. The causes of this phenomenon are not yet completely understood. We view this paper as the first step towards understanding the greater degree of bidirectional connectivity of skitter AS graphs.

In operational practice ASes are distinguished not only by connectivity or size but also by type, e.g., carriers, exchange points, local telcos, consumer ISPs (dial-up, DSL, cable), content providers or content distributors. ASes may enter into a multitude of contractual relationships, the most common of which are peers, customers, providers and mutual transit providers (siblings). More complicated arrangements such as partial transit restricted to a geographical region, or a combination of being a customer and partial transit provider to an upstream AS, are not uncommon. Heuristics involving an AS's in- and outdegree, the number of transited and originated prefixes, and even the magnitude of AS number itself can be used to derive educated guesses about AS relationships.

This paper investigates topological properties of AS connectivity in finer detail. We analyze differences between AS paths obtained from BGP routing tables and AS paths derived from traceroute paths. We also introduce a new type of AS relationship, *common ownership*, that reflects the complexities of real-world business relationships and practices.

2 Previous work

Data collections. BGP was defined by Yakov Rekhter and Tony Li in 1995 [1]. Analyzing the BGP routing table as a source of global Internet data was pioneered in mid-90s by Erik-Jan Bos at SURFnet, Tony Bates (then at Cisco) [5], and Geoff Huston [6] [7]. Huston's work analyzes trends in Internet connectivity, including the dynamics of AS degrees, AS path counts, and AS path lengths and presents them as a daily report. David Meyer's RouteViews project [8] has collected BGP tables at the University of Oregon since 1997. RIPE has collected both BGP updates and BGP tables since 1999. The importance of these pioneering work and the public availability of these data collections can hardly be overestimated.

CAIDA's Skitter project has collected traceroute data since 1998 [9]. AS connectivity is derived from traceroutes and published periodically as an AS core poster [10]. Construction of IP-level Internet maps was also done by Cheswick and Burch [11] who made their traceroutes publicly available in 1999-2001.

Analysis. Internet path properties were originally studied by Vern Paxson in his Ph.D. thesis [12]. This and later work by Paxson and colleagues focus considerable attention on handling noise in data, with careful exposition of sources and consequences of errors in various classifications, including with routing pathologies. Their study of stationarity of Internet path properties [13] concluded that many IP paths (70-80% of their data) remained unchanged for longer than one day, suggesting that our analysis of 9-hour-long skitter cycles is not significantly affected by routing dynamics.

Routing stability issues are always intertwined with properties of AS connectivity. Several studies [3] [2] have confirmed that a rich mesh of connections among ASes can result in an extreme (superexponential in the number of nodes in between) amount of routing traffic caused by a single update. These results underscore the need for greater insight into AS-level topology and constraints that it imposes on the propagation of BGP updates.

Broido and Claffy [14] [15] used systems of AS paths observed from multiple vantage points to classify prefixes into groups called *policy atoms*. Informally, each atom consists of prefixes that are commonly routed in a large portion of the Internet. The authors show that the use of atoms can reduce the size of routing tables by half. The work of Yehuda Afek *et al.* [16] establishes the presence of atoms in BGP updates.

Properties and limitations of AS-level Internet maps have been discussed in several papers. Faloutsos *et al.* [17] presented evidence that the node degree distribution for a BGP AS graph is close to a power function. AS hop distances are also studied in Huston's report [6] and in modeling of BGP update traffic [18]. Average AS path length and peering richness (entropy of outbound link use by routing) has remained relatively stable in 1999-2001, changes at individual ASes notwithstanding [19].

Lixin Gao [20] classified links in the BGP AS graph as customer, provider, peer and sibling connections. Tangmunarankit *et al.* [21] used this classification to assess hop count differences between router-level and

AS-level shortest paths versus shortest paths constrained by policies as inferred in [20]. A related paper [22] evaluates topology generators on the basis of reachability functions, resilience and distortion (tree-likeness) of the resulting graphs.

Research on IP-level maps of global Internet connectivity was described in [23] [24]. Router-level mapping was initiated with the Mercator project [25], the first serious attempt at identifying all IP addresses assigned to the interfaces of a single router using the alias probe heuristic of [26]. A further router identification technique was implemented in `iffinder`, a tool developed by Ken Keys at CAIDA [27].

Chang *et al.* tackled the problem of identifying which AS owns a router [28], introducing heuristics to fill in ASes for non-replying hops in the trace. The paper also discusses third-party addresses—addresses belonging to an AS in the return path that does not appear in the forward traceroute path. Our analysis [29] suggests that this is a negligible source of AS path incongruity.

Ratul Mahajan *et al.* [30] studied BGP misconfigurations, classifying origin misconfigurations into 'related origin', when a new origin is present in an observed AS path upstream or downstream from the old origin, and 'foreign origin', when no such connection can be observed. Their results suggest that some of the cases in our data in which the BGP AS path is a truncated version of the traceroute AS path, or vice versa, may derive from a misconfiguration.

Neil Spring *et al.* [31] discussed how to build a detailed map of an individual ISP's topology using their tool Rocketfuel which employs some 800 traceroute servers to gather as much topological information possible with the minimal amount of measurement. They use a rich set of heuristics for identifying same-router IP addresses, including DNS names, TTLs, IP ID field, and instances of rate limiting triggered by earlier probes. Their detailed analysis results in the discovery of seven times more links than skitter [10] in selected networks of several ISPs. They do not attempt to obtain an AS-level map.

Lisa Amini *et al.* [32] compared properties of traceroute and BGP AS paths. They found that the IP stacks of AIX, FreeBSD, Windows 2000, and the Cisco 7500 set the source address of ICMP reply messages to the interface on which the packet triggering the response arrived. They also found that only Linux sets it to the interface on which the reply is sent, as required by specification [33]. They also found a significant number of cases in which the the traceroute AS path contained exactly one more AS than the corresponding BGP AS path. They suggest that some, perhaps most, of these additional ASes are the ASes of exchange points. Elena Silenok [34] examined how much peering at exchange points could be observed in skitter paths.

Lisa Amini's work motivated us to expand earlier analyses of IP and AS graphs [4] [14] [19] [10] to better understand the relative richness of traceroute AS graphs compared to BGP AS graphs. For example, the core of a traceroute AS graph is 10 times larger than the core of a BGP AS graph, which suggests that the former is qualitatively superior. Our present paper contains the initial results of this investigation.

3 Methodology

3.1 Collection of traceroute and BGP AS paths

Our study is based on data collected at three locations worldwide. At each location we obtained traceroute IP paths and BGP AS paths from two topologically nearby hosts. The close proximity of each probing host to a corresponding routing table allows us to compare aligned views of the network from each vantage point.

3.1.1 Source of traceroute paths

A traceroute IP path is an approximation of the hop-by-hop router-level forward path a packet would take to a destination. There is currently no direct and failsafe method of obtaining forward IP paths. Instead, programs like `traceroute` [35] infer an IP path by sending out cleverly crafted probe packets and then analyzing the ICMP error responses. CAIDA's skitter monitors [9] employ a similar technique to determine the forward path to thousands of destinations. Each monitor continuously cycles through a destination list containing the IP addresses to probe. By overlapping the probes to multiple destinations at a time, skitter can probe a large number of destinations in a relatively short period of time—between a few hours to a single day for several hundred thousand destinations. The short duration of each cycle reduces the probability of distortions being introduced by changes in routing.

The traceroute IP paths of this study were obtained from three skitter monitors. The skitter monitor at San Jose, `sjc`, is located in the network of MFN/AboveNet (AS6461). The Amsterdam monitor, `k-peer`, is located in the network of RIPE (AS 3333) and lies near the Amsterdam Internet Exchange (AMS-IX). The Tokyo monitor, `m-root`, is located in the same network as the *m* root DNS server (AS 7500) and lies near the Network Service Provider Internet eXchange Project (NSPIXP) hosted by WIDE.

We use only a single cycle of traces from each monitor, with all traces taken between 1:00 and 13:00 (PST) on April 1, 2002.[1] During this period, `k-peer` and `m-root` were probing the same destination list (DNS), while `sjc` was probing a different list (IPv4). The DNS list contains 143,193 addresses and consists only of hosts that have queried one of several instrumented DNS root servers. The IPv4 list contains 301,752 addresses and consists of a wide variety of hosts, including dial-up and consumer broadband users. The two lists have 23,903 addresses in common.

Some trace attempts fail because of packet loss, ICMP filtering or rate limiting, unreachable destinations, transient routing problems, and other causes. We use only 'complete' traces—traces in which the destination and all intermediate hops replied. For `sjc`, 220,088 (73%) of 301,752 attempted traces are complete. For `k-peer` and `m-root`, 89,667 (63%) and 89,317 (62%) of 143,193 traces are complete, respectively.

3.1.2 Source of BGP paths

A BGP AS path describes the forward path at the level of ASes. BGP AS paths are an integral part of the BGP interdomain routing protocol and reflect the propagation of routing updates between interdomain routers participating in BGP. In the normal operation of BGP, ASes announce their address space to their neighbors in units of address prefixes. Before propagating them further, recipients of these routing updates prepend their AS number to the AS path enclosed in the update. In this way, as updates propagate across the Internet, the AS path associated with each announced prefix is extended to reflect the path taken by the updates.

In contrast to forward IP paths, there is no need to infer BGP AS paths. Every interdomain router has a BGP routing table that contains the AS path of every prefix it knows about. For our study, we use the BGP routing table of a backbone router located within the same AS and within a few IP hops from each skitter monitor.[2] These backbone routers are peers of RouteViews and thus export their routing table to the RouteViews collector. This allowed us to obtain these routing tables from the regular snapshots provided by RouteViews. We also chose these particular routers because they have complete routing tables—they have an entry for every announced prefix.[3]

We obtained BGP AS paths from the RouteViews snapshot taken at 6am (PST) on Apr 1, 2002. We selected this time because it is in the middle of the period spanned by our skitter traces. The backbone routers closest to our measurement stations at `sjc`, `k-peer`, and `m-root` have 107,707, 116,077, and 116,787 prefixes in their routing tables, respectively.

3.1.3 Conversion of IP paths to AS paths

We convert traceroute IP paths to AS paths by longest-prefix matching. We first determine the AS that originates (originally announces) each prefix. This is assumed to be the last AS of each BGP AS path (but note [30]). Then, for each IP address in an IP path, we find the longest announced prefix that contains the address. The origin AS of the matching prefix is the AS of the IP address.

[1]The `sjc`, `k-peer`, and `m-root` monitors took around 9, 7, and 7 hours, respectively, to complete the cycles used in this paper. This data is included in CAIDA's Internet Topology Data Kit (ITDK0204), available at `http://www.caida.org/tools/measurement/skitter/idkdata.xml`. ITDK was used in some recent work by researchers at IBM and Columbia University.

[2]For `sjc`, loopback0.mpr2.sjc6.us.mfnx.net (207.126.96.1); for `k-peer`, nikrtr.ripe.net (193.0.0.56); and for `m-root`, m-gw.nspixp2.wide.ad.jp (202.249.2.86).

[3]However, routers at different locations on the Internet may see slightly different sets of prefixes due to policy decisions or connectivity arrangements. Thus there is no single definitive Internet BGP routing table. See [19] for details.

Because of IP addresses for which no corresponding AS legitimately exists, it is not possible to fully convert every traceroute IP path to an AS path. Some addresses cannot be mapped to an AS because they lack a matching prefix in the RouteViews snapshot. Other addresses cannot be mapped because they fall within reserved address spaces, such as the RFC1918 private address space, that do not belong to any single organization.[4] We discard these unmappable addresses (but not the paths containing them) early in our analysis, during the conversion of IP paths to AS paths and during the determination of the initial set of incongruent AS paths (see Section 3.1.5 below). About half of the occurrences fit the pattern $\alpha\, X\, \alpha$, where X is the unmappable address and α is the AS of adjacent IP addresses. Because the adjacent addresses map to the same AS α, discarding the unmappable address is not likely to distort later analysis.

3.1.4 Pairing of traceroute and BGP AS paths

In our analysis, we compare traceroute AS paths to their corresponding BGP AS paths. Each traceroute path has a single natural corresponding BGP AS path—namely, the BGP AS path of the longest prefix that matches the destination address of each traceroute path.[5] The same longest-prefix matching is used internally by routers in forwarding each packet. We are thus comparing the presumed actual path with the announced path for each destination.

After pairing traceroute and BGP AS paths, we eliminate redundant pairs so that our statistics will not be distorted by overrepresented groups. A pair is redundant if there exists another pair *for the same BGP prefix* that has the same traceroute AS path. Redundant pairs occur because skitter destination lists typically include more than one destination in each prefix, leading to multiple traceroute IP paths per prefix, with many IP paths yielding the same AS path. The number of non-redundant pairs of AS paths for `sjc`, `k-peer`, and `m-root` is 60,271, 36,950, and 38,527, respectively.[6] These numbers represent 27%, 41%, and 43% of the complete skitter traces in each respective dataset. The corresponding number of BGP prefixes covered by these non-redundant pairs is 58,037, 36,170, and 37,292, representing 54%, 31%, and 32% of total announced prefixes. Around 97% of BGP prefixes are paired with just a single traceroute AS path after redundant pairs are eliminated. Given the percentages of BGP prefixes covered, we have some confidence of having obtained a representative sample of traceroute AS paths.

3.1.5 Initial set of incongruent AS paths

The initial set of incongruent AS paths is the departure point for the remaining analyses in this paper. We derive this set in a straightforward manner from the non-redundant pairs of traceroute and BGP AS paths. An AS path pair is incongruent if the traceroute and BGP AS paths have an unequal length or if the ASes at corresponding positions are not numerically equal.[7] The initial number of incongruent AS paths for `sjc`, `k-peer`, and `m-root` is 11,297, 36,888, and 38,460, respectively. The corresponding number of BGP prefixes covered by these paths is 10,519, 36,108, and 37,226.

---

[4]The three sets of traceroute IP paths have a varying number of these unmappable IP addresses, and the number of affected paths also varies. There were 7,667 such addresses for `sjc`; 19% have no matching prefix, 80% are RFC1918 or loopback addresses, and the remainder fall within other reserved spaces. About 8% of the complete paths have at least one unmappable address. `K-peer` and `m-root` have 934 and 3,829 unmappable IP addresses, respectively, and 3% and 7% of their complete paths have these addresses.

[5]A small number of destinations do not have a matching prefix and thus cannot be paired up in this way. There are 40, 31, and 12 such paths for `sjc`, `k-peer`, and `m-root`, respectively.

[6]When we convert a traceroute IP path to an AS path, we replace unmappable IP addresses with a special AS number. We eliminate these special ASes when determining the initial set of incongruent AS paths, but they still exist at this stage and cause us to slightly overestimate the number of non-redundant AS paths. The number of non-redundant pairs is 992 less for `sjc`, 37 less for `m-root`, and unchanged for `k-peer` when these special ASes are eliminated. Because these differences have only negligible effects on subsequent analysis, we continue to refer to the original numbers in the remainder of this paper.

[7]All unmappable IP addresses are eliminated from the non-redundant path pairs before comparisons are made. In most cases, we simply discard unmappable addresses, but sometimes we can infer their AS. If we can align the traceroute and BGP AS paths in such a way that some subsequence $\alpha\, X\, \beta$ of the traceroute AS path corresponds to the BGP subsequence $\alpha\, Y\, \beta$, then we infer the AS Y for the unmappable address X.

3.2 Compilation of exchange point ASes

Exchange points, network access points, and GigaPoP's (collectively called Internet exchange points, or IXes, in this paper) are peering points where network providers interconnect and exchange traffic destined for each other's customers. An IX is typically a location at which participating providers connect their routers to a common network. Routers at an IX have addresses taken from a prefix belonging to the IX. Passage through an IX can often be detected in traceroute IP paths by either matching addresses to known IX prefixes [36] or by examining the hostname at each hop. Occasionally a participating provider announces the IX prefix as its own.[8]

Because detecting passage through IXes is not the aim of our study, we do not make use of IX prefixes. We are only interested in such passage if it causes the AS path derived from the traceroute IP path to differ from the announced BGP AS path. Thus we rely instead on IX *ASes*. An IX AS is an AS that has been registered under the name of an IX. Examples are 6695 (DE-CIX), 1200 (AMS-IX), and 12536 (MaNAP). Not all IXes have ASes registered to them. Our list of IX ASes contains far fewer than the hundreds of IXes that exist in the world. An IX AS will appear in a traceroute AS path if a prefix announced by the IX AS best matches an address in the path.

IX ASes occur more often in traceroute AS paths than one would expect. They are the cause of many incongruities between traceroute and BGP AS paths (see Section 4.1). In fact, almost every occurrence of an IX AS in a traceroute path is a source of incongruity since such ASes rarely occur in BGP paths.

Since no ready-made list of IX ASes could be found, we compiled our own by querying the major Internet registries (ARIN, RIPE, APNIC) and a routing registry (RADB). Registry records are often outdated but usually in parts, such as contact addresses, that do not affect our analysis. Our list will be inaccurate only if an AS is reassigned to a completely unrelated new owner, which does not often occur. Our list contains 60 IX ASes[9]. We include AS 2500 (WIDE) in our list even though the owner itself is not an IX. WIDE hosts the Internet exchange NSPIXP, and NSPIXP is the source of nearly all occurrences of AS 2500 in one set of traceroute AS paths on Apr 1st. These occurrences are caused by routers with names like `as4732.nspixp2.wide.ad.jp`.

3.3 Compilation of AS ownership groups

In theory, no organization needs more than a single AS. In practice, many organizations have more than one, and some have dozens. There are diverse reasons for having multiple ASes, such as for convenience in implementing routing policy and for segregating different classes of traffic (e.g., academic vs. commercial). Businesses can also find themselves possessing multiple ASes after an acquisition or merger. Previous studies of AS connectivity do not take into account the complexities introduced by the common ownership of ASes. The realities of common ownership cause the definition of equality between two ASes to change from simple numerical equality to a test of membership in an equivalence class. As a result, statistics and relationships determined at the AS level may yield an incomplete picture if common ownership is not considered. For example, many researchers study peering relationships at the AS level. If an organization uses one AS to peer with one set of organizations and another AS to peer with a different set of organizations, then a true count of the number of peers cannot be determined without taking common ownership into account.

Compiling a list of AS ownership groups is challenging. For example, it requires a history of business acquisitions and mergers, which no one has assembled. Even finding all the ASes registered by a single organization (ignoring subsidiaries) is difficult. We employed several methods to compile a list of ownership groups. Our primary method was to correlate records from Internet registries. We had available a copy of a subset of the ARIN WHOIS database that provides the name and the database ID of the registered owner of each AS allocated by ARIN. We grouped records by database ID and by name. Because the name field is free-form text entered by the registrant, there is often slight variation in the text supplied by a single organization

[8]The announcement of IX prefixes by participating providers is a possible cause of some of the remaining incongruities discussed in Section 4.3.

[9]We include the following ASes that are designated for 'Exchange Point Blocks': 2886, 4551, 4552, 4554-4556, and 4558-4561.

for their different ASes (such as in the use of "Inc."). We sorted the records by the name field to put similar names near each other despite slight variations and then manually grouped the records. We also applied the same method to a file used by the CIDR Report [5] for mapping ASes to the names of organizations. This file appears to have been compiled automatically from all the major registries and seems to cover nearly all registered ASes.

We sought out press releases and news reports to confirm suspected business relationships. We found hints of these relationships in the traceroute IP paths themselves—in the hostnames and in the observed peering relationships between ASes. We also verified our groupings by manually querying the Internet registries such as RADB. A manual query can obtain the full record, which includes information on technical contacts, domain names, and registered routes. These clues are often sufficient to either confirm or refute a grouping.

4 Classification of incongruities

Based on the cause of discrepancy, we classify traceroute and BGP AS path incongruities into three categories. The first category consists of traceroute and BGP AS paths that differ from each other only in the appearance of one or more IX ASes. The second category consists of AS paths that differ because of mismatches between ASes under the same ownership. The third category consists of all remaining incongruities. Although this last category lumps together disparate causes, it does provide a rough upper bound on the magnitude of any single unknown cause of incongruities.

Unless otherwise noted, numbers and percentages in this section concerning AS paths refer to the total number of *non-redundant* traceroute AS paths (see Section 3.1 for a discussion of how this set differs from the set of all traceroute AS paths).

4.1 Incongruities arising from exchange point ASes

A large number of traceroute and BGP AS paths differ only in the appearance of IX ASes. Although IX ASes are rarely seen in BGP AS paths, they are not uncommon in traceroute AS paths.

Table 1 shows the number of occurrences of the top 5 most frequently occuring IX ASes in traceroute and BGP AS paths.[10] There are only a few distinct IX ASes in our dataset: 17 for `sjc`, 19 for `k-peer`, and 20 for `m-root`. Furthermore, only a few distinct IX ASes are responsible for the majority of occurrences; for example, just 4 IX ASes are responsible for over 90% of all occurrences of IX ASes in `sjc` AS paths. `K-peer` is located near the exchange points AMS-IX (AS 1200), and `m-root` is located near WIDE/NSPIXP (AS 2500). As a consequence, almost all their paths[11] traverse these exchange points, leading to the unusually large number of occurrences of these IX ASes. In contrast to traceroute AS paths, IX ASes are relatively rare in BGP AS paths.[12]

Table 2 shows the number of traceroute AS paths that include IX ASes. The first row gives the number of AS paths that have no IX AS; the second row gives the number of AS paths that have one IX AS; and so on. Few paths have more than one IX AS. In order to have more useful numbers in this table, we have excluded the occurrences of AMS-IX and WIDE/NSPIXP, the exchange points near `k-peer` and `m-root`, from the counts for `k-peer` and `m-root`, respectively.

Table 3 shows the breakdown of the initial set of incongruent traceroute and BGP AS paths in terms of IX ASes. Some paths differ only by the insertion/deletion of IX ASes (row 1), only by non-IX ASes (row 2), or by a mixture of IX and non-IX ASes (row 3). Thus IX ASes are responsible for causing 33%, 82%, and 54% of all AS path differences in `sjc`, `k-peer`, and `m-root` traces, respectively. The number of incongruent paths

[10] We only count those occurrences that cause an incongruity between traceroute and BGP AS paths.

[11] 99.8% for `k-peer` and 83% for `m-root`

[12] There are just 5, 4, and 7 distinct IX ASes in the BGP AS paths of `sjc`, `k-peer`, and `m-root`, respectively. These distinct IX ASes are responsible for fewer than 100, 350, and 150 occurrences of IX ASes in the BGP AS paths of non-redundant AS path pairs for `sjc`, `k-peer`, and `m-root`, respectively. The most common IX ASes seen in the BGP AS paths collectively are STAR TAP (AS10764), HKIX (AS 4635), SAIX (AS 5713), and HIX (AS 9035).

sjc			k-peer			m-root		
IX AS	freq	cum.%	IX AS	freq	cum.%	IX AS	freq	cum.%
6695	2,174	48.1%	1200	36,908	98.0%	2500	31,679	90.9%
5459	1,187	74.3%	10764	305	98.8%	7527	1,423	95.0%
7527	546	86.4%	6695	252	99.5%	6695	949	97.7%
2500	176	90.3%	8235	49	99.6%	5459	388	98.8%
1120	164	93.9%	5713	32	99.7%	1120	91	99.1%
total	4,331		total	36,150		total	33,782	

Table 1: Top 5 most frequently occurring IX ASes in non-redundant AS paths (# occurrences of IX ASes).

# IX ASes per path	sjc		k-peer		m-root	
0	54,546	(91%)	35,880	(97%)	34,329	(89%)
1	5,648	(9%)	1,052	(3%)	4,060	(11%)
2	77	(0%)	18	(0%)	118	(0%)
3	0	(0%)	0	(0%)	20	(0%)
total	60,271	(100%)	36,950	(100%)	38,527	(100%)

Table 2: Distribution of IX ASes in non-redundant AS paths [traceroute AS paths only] (# paths).

caused by IX ASes is higher for `k-peer` and `m-root` than for `sjc`, since most paths originating at `k-peer` and `m-root` include the AS of the IX near these measurement hosts.

Cause of Incongruity	sjc		k-peer		m-root	
only IX ASes	3,749	(33%)	30,163	(82%)	20,601	(54%)
only non-IX ASes	6,818	(60%)	4	(0%)	6,759	(18%)
both IX & non-IX ASes	712	(6%)	6,721	(18%)	11,100	(29%)
total incongruent paths	11,279	(100%)	36,888	(100%)	38,460	(100%)

Table 3: Breakdown of the initial set of incongruent AS paths by cause (# paths).

IX ASes are a significant cause of incongruity between traceroute and BGP AS paths. Because their appearance affects the adjacency of ASes in AS paths derived from traceroute IP paths, they can distort analyses unless dealt with. For example, AS links introduced by IX ASes can cause spurious cycles in derived AS graphs. We speculate that this may be one of the reasons why the core of a traceroute AS graph is 10 times larger than the core of a BGP AS graph.

Nevertheless, incongruities caused by IX ASes are largely superficial. In most cases, the same sequence of intermediate providers announced via BGP is traversed in traceroute, and thus the traceroute and BGP AS paths agree on essential points. There are cases, however, in which the traceroute path appears to represent a shortcut through an IX. We leave it to future work to determine the frequency and the cause of such apparent shortcuts.

Exchange points can be the source of additional incongruities. ISPs participating at IXes can sometimes (erroneously) announce IX prefixes themselves. This can cause IX addresses in traceroute paths to be mapped to these participating ASes, leading to an incongruity if the preceding or following hops are in a different AS. These spurious incongruities are not treated separately in this study. They simply contribute to the number of unidentified incongruities in the following analyses.

4.2 Incongruities arising from ASes under the same ownership

ASes under the same ownership are responsible for causing the apparent incongruity between a number of traceroute and BGP AS paths. For `sjc`, of the 7,530 AS paths that differ after IX ASes are removed, 2,711 (36%) paths differ due to the insertion/deletion of ASes under the same ownership. The corresponding numbers for `k-peer` and `m-root` are 1,464 (22%) of 6,725 remaining AS paths and 932 (5%) of 17,859 remaining AS paths, respectively.

Table 4 shows the number of occurrences of the top 10 most frequently occurring AS ownership groups in non-redundant AS paths.[13] The total number of distinct ownership groups seen at `sjc`, `k-peer`, and `m-root` is 99, 76, and 87, respectively. However, just a few ownership groups are responsible for the majority of occurrences. Fifty percent of all occurrences are attributable to just 4, 1, and 4 ownership groups for `sjc`, `k-peer`, and `m-root`, respectively. Similarly, 90% of all occurrences are attributable to 17, 11, 18 groups, respectively.

Many of the greatest contributors are large ISPs or transit providers. This is not surprising, since such organizations tend to grow by acquisitions and mergers, and at least initially, they must keep the merged networks, each with its own AS number(s), distinct. Providers also tend to have networks covering different geographical scales (such as regional, national, and international) and networks targeted to different classes of customers (e.g., academic vs. commercial and transit vs. dial-up). Each of these dimensions might use different AS numbers.

sjc			**k-peer**			**m-root**		
group	freq	cum.%	group	freq	cum.%	group	freq	cum.%
MCI[a]	665	16%	Level3	3,518	53%	MCI	1,719	33%
SBC[b]	571	29%	C&W[c]	465	60%	Telia	444	41%
Qwest[d]	557	42%	Telia	329	65%	Qwest	377	48%
Telia	494	54%	Qwest	322	70%	SBC	369	55%
AT&T	323	62%	XO	300	75%	Sprint	355	62%
Sonera	230	67%	SBC	261	78%	AT&T	283	67%
Italia[e]	196	72%	MCI	190	81%	Cogent	188	71%
Cogent [f]	166	76%	Cogent	151	84%	Italia	162	74%
CSU[g]	159	79%	Italia	132	86%	COLT	128	76%
ImpSat	107	82%	AT&T	128	88%	Level3	125	79%
total	4,232		total	6,623		total	5,284	

[a]MCI/WorldCom/UUNET/AlterNet/ANS/Bertelsmanns
[b]SBC/Pacific Bell/Nevada Bell/Southwestern Bell
[c]C&W/Exodus/PSI
[d]Qwest/US West/SuperNet/Touch America
[e]Telecom Italia
[f]Cogent/PSINet/NetRail
[g]California State University

Table 4: Top 10 most frequently occurring AS ownership groups in non-redundant AS paths (# occurrences of groups).

AS ownership groups arise from the unavoidable complexities of real-world relationships. As our data shows, these complexities intrude upon both traceroute and BGP AS paths more frequently than many realize. Much prior research examines ASes in isolation, inferring, for example, the peering relationship between two *ASes* when what is really sought is the peering relationship between two *organizations*. Such analyses suffer from inaccuracies because of the commonly accepted simplifying assumption that ASes can be treated separately

[13]For `sjc`, `k-peer`, and `m-root`, respectively, the total number of occurrences of ownership groups in (a) traceroute AS paths is 2,758, 3,610, and 3,084 and (b) BGP AS paths is 1,474, 3,013, and 2,200.

without regard to real-world relationships.

4.3 Remaining incongruities

A significant number of incongruent paths remain beyond those that differ by IX ASes or AS ownership groups: 4,819 at `sjc`, 5,216 at `k-peer`, and 16,927 at `m-root`. These numbers represent 43%, 14%, and 44% of the initial incongruent AS paths and 8%, 14%, and 19% of the non-redundant AS paths. The following textual analysis of the remaining incongruent AS paths provides some insight into the causes, although a precise classification of the causes is left for future work. We perform the analysis on the original AS paths without discarding IX ASes or collapsing AS ownership groups.

Table 5 shows the difference in length ($t - b$) between the remaining incongruent traceroute (t) and corresponding BGP (b) AS paths.[14] Most traceroute AS paths are longer than their BGP counterpart; the exact counts appear at the end of the table in the row labeled +. In the case of `k-peer` and `m-root`, few BGP AS paths are longer than the corresponding traceroute paths (the counts appear in the row labeled −), due primarily to the inclusion of the AS of the IX located near them in the traceroute paths.[15] We speculate that some of the increase in length is due to the appearance of customer ASes at the tail of the traceroute AS paths.

sjc			**k-peer**			**m-root**		
$t-b$	freq		$t-b$	freq		$t-b$	freq	
2	1,597	(33%)	2	1,861	(35%)	2	7,795	(46%)
-1	1,203	(25%)	0	1,485	(28%)	1	6,217	(37%)
1	1,151	(24%)	3	941	(18%)	3	1,220	(7%)
0	474	(10%)	1	679	(13%)	0	1,126	(7%)
3	315	(7%)	4	148	(3%)	4	468	(3%)
4	53	(1%)	-1	98	(2%)	5	57	(0%)
*	26	(1%)	5	36	(1%)	-1	25	(0%)
			*	13	(0%)	*	19	(0%)
+	3,125	(65%)	+	3,673	(70%)	+	15,765	(93%)
−	1,220	(25%)	−	103	(2%)	−	36	(0%)
total	4,819		total	5,216		total	16,927	

Table 5: Difference in length ($t - b$) between the remaining incongruent traceroute (t) and BGP (b) AS paths (# paths).

We examined the differences between corresponding AS paths and assigned each pair a metric derived from the edit distance between them. For each pair of incongruent AS paths we compute the minimal set of editing operations needed to transform the BGP AS path into the traceroute AS path. We use the UNIX `/bin/diff` program and allow insertions, deletions and substitutions. Specifically, the editing operations are restricted to: (1) *insertion* of ASes into traceroute AS paths, (2) *deletion* of ASes from BGP AS paths, and (3) *substitution* of a block of ASes in the traceroute AS path for a block of ASes in the BGP AS path, with the blocks having no ASes in common (that is, the blocks are disjoint sets of ASes). A block is a subsequence of ASes in an AS path; the two blocks involved in a substitution need not have the same length. We call this process of determining the edit distance an *aligning* of AS paths.

For example, in the following pair of incongruent AS paths, the alignment editing operations are 'delete 11422' and 'insert 1':

[14] Entries with an asterisk in the $t - b$ column indicate the number of AS paths that have length differences other than those included in the table.

[15] However, 2,994 (57%) and 9,548 (56%) traceroute paths for `k-peer` and `m-root`, respectively, are longer by at least 2 ASes, and the insertion of a single IX AS would not fully explain their increase in length.

```
BGP          207.99.128.0/17    6461 209 11422 2151   2920
Traceroute   207.99.161.1       6461 209       2151 1 2920
```

In the following pair, the editing operation is 'substitute 852 for 6453 271':

```
BGP          142.32.0.0/14      6461 6453 271 3633
Traceroute   142.32.1.1         6461      852 3633
```

Table 6 shows the breakdown of editing operations needed to align corresponding BGP and traceroute AS paths. For the majority of paths, the traceroute AS path can be derived from the BGP AS path by merely inserting some ASes. Alignment by merely deleting ASes is practically nonexistent for `k-peer` and `m-root`; however for `sjc`, deletions are a significant fraction of the required edits. `K-peer` and `m-root` have more alignments requiring substitutions and more complex derivations involving a sequence of editing operations. These observations suggest that the traceroute and BGP AS paths of `k-peer` and `m-root` differ more radically or more fundamentally than those of `sjc`.

Operation	**sjc**		**k-peer**		**m-root**	
insertions only	2,788	(58%)	2,764	(53%)	13,661	(81%)
deletions only	1,132	(23%)	1	(0%)	0	(0%)
substitutions only	813	(17%)	1,813	(34%)	2,648	(16%)
mixture	86	(2%)	683	(13%)	618	(4%)
total paths	4,819	(100%)	5,216	(100%)	16,927	(100%)

Table 6: Breakdown of remaining incongruent AS paths by editing operations needed to transform BGP AS paths to traceroute AS paths (# paths).

For `sjc`, 1,357 traceroute AS paths (12% of the initial number of incongruent paths from `sjc`) are simply extensions of the corresponding BGP AS paths; that is, they differ by insertions only and all insertions occur at the end. Some of these additional ASes at the end are likely customer ASes that are hidden in the BGP paths because of prefix aggregation. For `m-root`, only 2 traceroute AS paths are extensions, and `k-peer` has none. The near absence of path extensions for `m-root` and `k-peer` may be explained by their use of the DNS destination list. The DNS list consists of hosts that have queried a root DNS server, and these hosts are likely nameservers or other infrastructure located in ISPs and not at customer premises. Hence, aggregation of customer networks would occur less frequently for these datasets.

Some traceroute and BGP paths differ in every AS except the first (the AS of the measurement host) and the last (the AS of the destination). The number of such paths for `sjc`, `k-peer`, and `m-root` are 563, 233, and 251, respectively.

Table 7 shows a breakdown of the number of ASes involved in substitutions (# BGP ASes$\rightarrow$# traceroute ASes). Most substitutions replace a single AS with another. Because we compute a *minimal* set of editing operations, this means that the two ASes participating in a $1 \rightarrow 1$ substitution have the same preceding and following ASes in their respective AS paths. That is, suppose X and Y are ASes involved in a $1 \rightarrow 1$ substitution. Then some BGP path segment $\alpha\, X\, \beta$ becomes the traceroute segment $\alpha\, Y\, \beta$, with α and β being the same for the two paths.[16] This implies that X and Y both peer with α and β. This similarity in peering relationships suggests some possible causes. For example, X and Y could be the immediate providers of a multi-homed customer, and although X appears in the announced path, actual traffic crosses Y. Or, X and Y could be ASes under the same ownership, and the organization prefers to use, or has standardized upon, X in its announcements while Y reflects the actual subnetwork of the organization that traffic is crossing.

[16]There will always be a preceding and a following AS since no substitutions can occur at the first or the last AS in either path. Both paths must include the AS of the measurement host and the AS of the destination, and these ASes cannot vary between the two paths.

	sjc		k-peer		m-root	
1→1	426	(52%)	1,157	(64%)	1,884	(71%)
1→2	208	(26%)	399	(22%)	315	(12%)
2→1	61	(8%)	3	(0%)	47	(2%)
2→2	29	(4%)	51	(3%)	156	(6%)
$* \to *$	89	(11%)	203	(11%)	246	(9%)
total	813	(100%)	1,813	(100%)	2,648	(100%)

Table 7: Breakdown of substitutions by # ASes involved: # BGP ASes →# traceroute ASes (# substitutions).

5 Conclusions

Using a comprehensive set of traceroutes from three locations worldwide, we taxonomized differences between actual and BGP-announced AS paths. We found two significant causes of disparity: (1) ASes assigned to exchange points, which rarely appear in BGP paths but often appear in forward IP paths, and (2) groups of ASes under the same ownership. Our results suggest that the observed difference in size between the cores of an AS graph derived from BGP and an AS graph derived from traceroute is due to the appearance of exchange point ASes in AS paths derived from traceroute paths.

We have established and employed a new type of AS relationship, *common ownership*, which reflects the complexities of real-world business relationships (arising, for example, from acquisitions and mergers) and practices (such as the use of different AS numbers for different classes of traffic).

The present paper describes a work-in-progress. In the future, we will employ BGP updates in determining incongruities and augment the analysis of IX ASes with IX prefixes in order to detect incongruities caused by providers announcing exchange point prefixes.

Acknowledgements

We would like to thank our colleagues at CAIDA for discussions and aid in collecting data. We are also thankful to the SIGMETRICS reviewers for comments that helped to improve this paper.

References

[1] Y. Rekhter and T. Li., "A Border Gateway Protocol 4 (BGP-4), RFC1771," Mar 1995.

[2] Tim Griffin, "What is the sound of one route flapping?," http://www.cs.dartmouth.edu:80/%7Emili/workshop2002/slides/griffin_dart%mouth_20020723.pdf.

[3] Craig Labovitz, Abha Ahuja, Roger Wattenhofer, and Srinivasan Venkatachary, "The Impact of Internet Policy and Topology on Delayed Routing Convergence," Apr 2001, Proceedings of INFOCOM 2001, Anchorage, AK.

[4] Andre Broido and k claffy, "Internet Topology: connectivity of IP graphs," in *Scalability and Traffic Control in IP Networks*, Aug 2001, Proceedings of SPIE, vol.4526.

[5] Tony Bates, "CIDR Report," http://www.cidr-report.org.

[6] Geoff Huston, "BGP Table Data (daily reports)," http://bgp.potaroo.net.

[7] Geoff Huston, "Analyzing the Internet's BGP Routing Table," *The Internet Protocol Journal*, vol. 4, Mar 2001, http://www.telstra.net/gih/papers/ipj/4-1-bgp.pdf.

[8] David Meyer, "University of Oregon Route Views Project," http://www.routeviews.org.

[9] Daniel McRobb and kc claffy, "Skitter," CAIDA, 1998. http://www.caida.org/tools/measurement/skitter/.

[10] B. Huffaker, A. Broido, k. claffy, M. Fomenkov, K. Keys, Y.Hyun, and D. Moore, "Skitter AS Internet Graph," Apr 2002, http://www.caida.org/analysis/topology/as_core_network/.

[11] B.Cheswick and H.Burch, "Internet Mapping Project," http://cm.bell-labs.com/who/ches/map.

[12] Vern Paxson, "End-to-End Routing Behavior in the Internet," in *IEEE/ACM Transactions on Networking*, Oct 1997, vol. 5, pp. 601–615.

[13] Y. Zhang, V.Paxson, and S.Shenker, "The Stationarity of Internet Path Properties: Routing, Loss and Throughput," in *ACIRI Technical Report*, May 2000.

[14] A. Broido and k claffy, "Complexity of global routing policies," http://www.caida.org/outreach/papers/2001/CGR/.

[15] A. Broido and k claffy, "Analysis of Route Views BGP data: policy atoms," Proceedings of network-related data management (NRDM) workshop, Santa Barbara, May 2001.

[16] Yehuda Afek, Omer Ben-Shalom, and Anat Bremler-Barr, "On the structure and application of BGP policy Atoms," in *Proceedings of the IMW workshop 2002. Marseille, France*, Nov 2002.

[17] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, "On power-law relationships of the Internet topology," Aug 1999, Proceedings of ACM SIGCOMM 1999, Cambridge, MA.

[18] Olaf Maennel and Anja Feldmann, "Realistic BGP Traffic for Test Labs," in *ACM SIGCOMM 2002, Pittsburg, PA*, Aug 2002.

[19] A. Broido, E. Nemeth, and kc claffy, "Internet expansion, refinement and churn," European Transactions on Telecommunications, 13, No.1, Jan-Feb 2002, 33-51.

[20] Lixin Gao, "Inferring Autonomous System Relationships in the Internet," Nov 2000, Proceedings of IEEE Global Internet Symposium, San Francisco, CA.

[21] H.Tangmunarankit, R.Govindan, and S.Shenker, "Internet path inflation due to policy routing," in *Scalability and Traffic Control in IP Networks*, Aug 2001, Proceedings of SPIE, vol.4526.

[22] H.Tangmunarunkit, R.Govindan, S.Jamin, S.Shenker, and W.Willinger, "Network Topology Generators: Degree based vs. Structural," in *Proceedings of ACM SIGCOMM 2002*.

[23] Hal Burch and Bill Cheswick, "Mapping the Internet," in *IEEE Computer*, Apr 1999, vol. 32.

[24] k. claffy, T. Monk, and D. McRobb, "Internet Tomography," Nature Magazine, Web Matters. http://helix.nature.com/webmatters/tomog/tomog.html.

[25] Ramesh Govindan and Hongsuda Tangmunarunkit, "Heuristics for Internet Map Discovery," in *Proceedings of IEEE Infocom 2000, Tel Aviv, Israel*, 2000.

[26] J.-J.Pansiot and D.Grad, "On Routes and Multicast Trees in the Internet," in *ACM SIGCOMM Computer Communication Review*, Jan 1998, vol. 28.

[27] Ken Keys, "iffinder," Feb 2001, http://www.caida.org/tools/measurement/iffinder/.

[28] H.Chang, S.Jamin, and W.Willinger, "Inferring AS-level Internet topology from router-level path traces," in *Scalability and traffic control in IP networks*, Aug 2001, Proceeding of SPIE, vol.4526.

[29] Young Hyun, Andre Broido, and kc claffy, "On Third-party Addresses in Traceroute Paths," Apr 2003, Passive and Active Measurement Workshop 2003, La Jolla, CA.

[30] R.Mahajan, D.Wetherall, and T.Anderson, "Understanding BGP Misconfiguration," in *ACM SIGCOMM 2002, Pittsburg, PA*, Aug 2002.

[31] N. Spring, R. Mahajan, and D. Wetherall., "Measuring ISP Topologies with Rocketfuel," in *ACM SIGCOMM 2002, Pittsburg, PA*, Aug 2002.

[32] Lisa D. Amini, Anees Shaikh, and Henning G. Schulzrinne, "Issues with inferring Internet topological attributes," in *Internet Performance and Control of Network Systems III. Proceedings of SPIE, vol.4865, Boston, MA*, Aug 2002.

[33] R.Baker, "Requirements for IP Version 4 Routers, RFC1812," Jun 1995.

[34] Elena Silenok, "Exchange Points as seen by skitter," `http://www.caida.org/~elena/projects/IX.html`.

[35] Van Jacobson, "traceroute," `ftp://ftp.ee.lbl.gov/traceroute.tar.Z`.

[36] Sean McCreary and Bill Woodcock, "Exchange Points," `http://www.pch.net/resources/data/exchange-points`.