Investigating Interdomain Routing Policies in the Wild

Ruwaifa Anwar
Stony Brook University
manwar@cs.stonybrook.edu

Haseeb Niaz
Stony Brook University
mniaz@cs.stonybrook.edu

David Choffnes
Northeastern University
choffnes@ccs.neu.edu

Ítalo Cunha
Universidade Federal de
Minas Gerais
cunha@dcc.ufmg.br

Phillipa Gill
Stony Brook University
phillipa@cs.stonybrook.edu

Ethan Katz-Bassett
University of
Southern California
ethan.kb@usc.edu

Abstract

Models of Internet routing are critical for studies of Internet security, reliability and evolution, which often rely on simulations of the Internet's routing system. Accurate models are difficult to build and suffer from a dearth of ground truth data, as ISPs often treat their connectivity and routing policies as trade secrets. In this environment, researchers rely on a number of simplifying assumptions and models proposed over a decade ago, which are widely criticized for their inability to capture routing policies employed in practice.

In this study we put Internet topologies and models under the microscope to understand where they fail to capture real routing behavior. We measure data plane paths from thousands of vantage points, located in eyeball networks around the globe, and find that between 14-35% of routing decisions are not explained by existing models. We then investigate these cases, and identify root causes such as selective prefix announcement, misclassification of undersea cables, and geographic constraints. Our work highlights the need for models that address such cases, and motivates the need for further investigation of evolving Internet connectivity.

Categories and Subject Descriptors

C.2.2 [Computer-Communication Networks]: Network Protocols

Keywords

Network Measurement; BGP; routing

1. INTRODUCTION

Research on existing and new protocols on the Internet is challenging because key aspects of the network topology are hidden from public view by interdomain routing protocols. Further, deploying new protocols at Internet scale requires convincing large numbers of autonomous networks to participate. As a result, networking researchers rely on assumptions, models, and simulations to evaluate new protocols [13,26], network reliability [20,41], and security [1,16,24].

Our existing models of interdomain routing [11], however, have important limitations. They are built and validated on the same incomplete topology datasets, typically routes observed via route monitors such as RouteViews and RIS [33, 39]. These vantage points expose a large fraction of paths from global research & education networks (GREN) and core networks, but they are incomplete in two keys ways. First, they expose few paths to and from eyeball and content networks. Second, they do not expose less preferred paths that would be used if the most preferred path was not available. As a result, they do not capture partial peering, more complex routing policies based on traffic engineering, or load balancing and the rich peering mesh which exists near the edge of the network [35].

While limitations of our existing models are well known [27,29, 35]–and are even being addressed in recent work [15]–we lack a solid understanding of how much these limitations impact our ability to accurately model the interdomain routing system. Recent work has attempted to address this issue by observing destination-based routing violations in control plane data [28] and by surveying a population of network operators about their policies [12]. However, these approaches are limited in terms of scale and their ability to observe behavior at the network edge.

In this paper, we take a systematic approach to understand how our models of routing policies [11] hold in practice. We leverage a combination of data plane measurements covering the network edge (Section 3.1) and control plane experiments which allow us to directly measure relative preference of routes (Section 3.2). We create a methodology that accounts for numerous potential causes of violations to our assumptions including sibling ASes [4], complex AS relationships [15], prefix-specific routing policies, and the impact of geography. We investigate the prevalence of each of these causes in AS-level paths observed via measurements of the data and control planes.

We revisit generally held assumptions and models of Internet routing. Our goal is *not* to measure a complete Internet topology; rather, we seek to improve our understanding of routing decisions made by ASes when routing their traffic. Towards this goal we make the following observations for our measured paths:

- Known hybrid and partial transit relationships (*e.g.*, those explored in [15]) contribute a surprisingly small amount to unexpected routing decisions.
- Per-prefix routing policies appear to explain 10-20% of unexpected routing decisions, where an AS chooses a longer or more expensive path than our model predicts.
- We find that some large content providers like Akamai and Netflix are destinations for a large fraction of unexpected routing decisions (21% and 17%, respectively).

IMC '15 October 28 - 30, 2015, Tokyo, Japan
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3848-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2815675.2815712.

- Routing decisions vary based on geography. We find that paths traversing multiple continents deviate from our models more, owing to undersea cable ASes which are not accounted for in our models. We also observed a tendency for ASes to prefer non-international paths when endpoints are in the same country.

Our results highlight areas where more investigation would yield the largest payoff in terms of improving our accuracy when modeling AS relationships and routing policies. We also identify key areas, specifically investigating prefix-specific routing policies, where additional vantage points and looking glass servers could improve the fidelity of our AS topology data.

2. MODELING INTERDOMAIN ROUTING

The now standard model of routing policies was developed by Gao and Rexford [10, 11] based on seminal work by Griffin, Sheppard, and Wilfong [17] and Huston [18, 19]. In this model, ASes connect to each other based on business relationships: (1) customer-provider, where the customer pays the provider, and (2) peer-to-peer, where the ASes exchange traffic at no cost. This model gives the following view of local preferences and export policies, based on the economic considerations of ASes:

Local Preferences. An AS will prefer routes through a neighboring customer, then routes through a neighboring peer, and then routes through a provider. In other words, an AS will prefer cheaper routes.

Export Policy. A customer route may be exported to all neighboring ASes. A peer or provider route may only be exported to customers.

This model is sometimes augmented with the assumption that ASes only consider the next hop AS on the path when making their routing decisions. This simplifies analysis and makes debugging more tractable [20]. Simulation studies also often restrict path selection to the shortest among all paths satisfying Local Preference and use tie-breakers to induce unique routing decisions when AS path lengths are same [13, 14].

While the above model and variations thereof have been used in many studies (*e.g.*, [1, 13, 16, 21, 41]), it is well known that this model fails to capture many aspects of the interdomain routing system [27, 29, 35]. These aspects include AS relationships that vary based on the geographic region [15] or destination prefix, and traffic engineering via hot-potato routing or load balancing.

Prior work has used traceroute measurements and BGP data to address some of these issues (*e.g.*, [27, 29]); however, these measurements only offer a glimpse into ASes' routing preferences. Namely, they expose only the set of paths that are in use at the time of measurements. In contrast, we use active control plane experiments (PEERING [37]) to expose less preferred paths. Further, these datasets have poor or no coverage of paths used by edge networks [7]. On a smaller scale, network operators were surveyed about their routing policies to better understand how our models correspond to practice [12], but the scale and representativeness of a survey approach makes generalizing these observations infeasible.

3. METHODOLOGY

We aim to understand the gap between interdomain routing models and empirically observed behavior on the Internet. Our methodology combines two measurement techniques to gain better visibility into interdomain routing policies. First, we passively observe routing decisions on paths towards popular content networks (Section 3.1). We leverage the RIPE Atlas platform which provides a large collection of vantage points located around the world for our traceroute measurements. We thus observe routing decisions for broad range of hosts from variety of vantage points. One limitation of this approach lies in its passiveness as it only provides information about paths that are in use at the time of measurements. We do not get any information about the alternate paths available to an AS. Our second technique (Section 3.2) overcomes the above mentioned limitation and exposes less preferred paths for different ASes. We use PEERING [2, 37, 40] to selectively poison BGP announcements and force ASes to choose an alternate path, then we use RIPE Atlas probes as vantage points to run traceroutes towards poisoned prefixes to observe these alternate paths. This approach of actively probing routing decisions enables us to discover less preferred paths and also reverse engineer the BGP decision process. However, the PEERING platform is currently limited to few locations from which we can send poisoned announcements.

3.1 Passively observing route decisions

It is well known that a disproportionately large amount of Internet traffic originates from a few popular content providers [23, 36]. However, there is little empirical data about the paths this traffic takes [23]. We target these paths with our measurements. Note that it is not our goal to observe routing decisions for the entire Internet. Rather, we focus on the more tractable task of measuring a subset of important Internet paths (those carrying most traffic) from a diverse set of vantage points, and putting those paths under the microscope to understand how and why they differ from paths predicted by routing models.

Selecting content providers. We consider a list of the top applications from Sandvine [36] and top Web sites from Quantcast [31]. From these lists, we isolate top HTTP and non-HTTP hosts in terms of number of downstream bytes and number of visits. Finally, we arrive at a list of 34 DNS names representing 14 large content providers.

AS type	Probes	Distinct ASes	Distinct Countries
Stub-AS	787	333	106
Small ISP	581	188	78
Large ISP	56	109	51
Tier 1	69	8	3

Table 1: Distribution of selected RIPE Atlas probes.

Vantage points (VPs). RIPE Atlas has broad global coverage, but is known to have a disproportionate fraction of probes skewed towards Europe. To avoid a bias towards European ASes, we picked equal number of probes from each continent. For every continent, we picked probes in a round robin fashion from different countries and ASes so that selected probes cover a wide range of ASes. Table 1 summarizes the location of these probes in terms of AS type using the categorization method of Oliveira *et al.* [30]. The bulk of the probes are located near the network edge in stub and small ISP networks. To measure paths to content providers, each RIPE Atlas node performs a DNS lookup for each of the 34 content DNS names, and then performs a traceroute to the resolved IP. We use 1,998 RIPE Atlas probes located in 633 ASes, distributed according to our sampling methodology.

Data set. We used maximum probing rate allowed by RIPE Atlas to perform 28,051 traceroutes towards selected hosts. These traceroutes ended up in a total of 218 destination ASes. The number of destination ASes is large relative the number of content providers because large numbers of content servers are hosted outside the provider's network (*e.g.*, inside ISPs) [5]. We convert the traceroute-based IP-level paths into AS paths using the method described by Chen et al. [7]. Since interdomain routing is destination-based, we can observe routing decisions for all ASes along the path to a given destination. We thus observe routing decisions for a total of 746 ASes.

3.2 Actively probing route decisions

Passive measurements observe only the most preferred route for an AS toward a destination. We use PEERING [2, 37, 40] to expose alternate, less preferred routes and to attempt to reverse engineer BGP decisions.

PEERING operates an ASN and owns IP address space that we can announce via several upstream providers. PEERING allows us to manipulate BGP announcements of its IP prefixes and observe how ASes on the path react. We used PEERING to announce prefixes using six US universities (Georgia Tech, Clemson, University of Southern California, Northeastern, Stony Brook, and Cornell) and one Brazilian university as providers. We change announcements at most once per 90 minutes to allow for route convergence and avoid route flap dampening. We use prefixes allocated to the PEERING research testbed reserved for our experiments; these prefixes carry no real traffic beyond our measurements.

Discovering alternate routes. We start announcing an IP prefix from all PEERING locations in an anycast announcement. At each round, we observe the preferred route at a target AS T and the next-hop neighbor N that T is using to route toward our prefix. We then poison N, *i.e.*, add N's AS number to the path [3, 9], to trigger BGP loop prevention at N and cause N to no longer have a path to our prefix (and stop announcing a route to T). This forces T to choose a different route, through a different neighbor N'. We repeat this process in consecutive rounds, poisoning the newly-discovered neighbor, to identify all neighbors and routes T can use toward our prefixes. When we observe different routes at the target AS T (through different neighbors) from multiple vantage points (*e.g.*, due to different routing preferences at different geographic locations), we run the algorithm for each vantage point separately. We can potentially execute this algorithm for each AS in the topology as the target AS. A similar experiment was performed by Colitti [9]; here, we use the same mechanism with a more diverse set of providers and with a different goal.

We insert all poisoned ASes into a single AS-set, and surround the poisoned AS-set with PEERING's AS number. This limits AS-path length, prevents inference of non-existent inter-AS links, and allows operators to identify the poisoning.

Reverse engineering BGP decisions. In addition to the experiment to discover alternate routes, we conduct a complementary experiment to infer BGP decision triggers. We first announce an IP prefix from one PEERING location (called the *magnet*), wait five minutes to allow for route convergence, then announce (*anycast*) the same IP prefix from all other PEERING locations. After we anycast the prefix, an AS may change to a new route with higher Local-Pref, shorter AS-path length, or better intradomain tie-breakers, as specified in the BGP decision process [8]. If an AS x keeps using the route toward the magnet after we anycast the prefix, we check if the magnet route is cheaper according to the Gao-Rexford model or has shorter AS-path length than all other routes we observed from x. If none of these checks are satisfied, we infer AS x is using intradomain costs or route age (the last tie-breaker before router ID) as a tie-breaker. If AS x did not choose the route to the magnet, we check if the chosen route is cheaper or shorter than the route to the magnet. If none of these checks are satisfied, we infer AS x is using intradomain costs as a tie-breaker.

We repeat this process using each PEERING location as the magnet. We also check whether the route chosen after we anycast the prefix is more expensive according to the Gao-Rexford model or is the same cost but has longer AS-path length than other routes we observed, which is a violation of the model. The route to the magnet may become unavailable when a downstream AS receives and chooses a more preferred route; in these cases we consider the downstream AS's decision.

Vantage points (VPs). We perform traceroutes from 96 RIPE Atlas probes and approximately 200 PlanetLab nodes every 20 minutes, and collect BGP feeds every 15 minutes from RouteViews and RIPE RIS to monitor paths toward PEERING prefixes. We use the maximum number of RIPE Atlas probes allowed within daily probing budget limits. We implement a greedy heuristic that picks probes to maximize the number of ASes traversed on the default paths toward PEERING locations.

Data set. We needed a total of 188 distinct poisoned announcements to infer preferences for all 360 target ASes we observe on paths toward PEERING (some poisonings are useful for multiple target ASes). We observe 739 inter-AS links. We find 45 inter-AS links that are not in CAIDA's AS-relationship database, 10 of which (22.2%) can only be observed with poisoned announcements.

3.3 Comparison with existing models

We compare paths observed in our passive and active measurements with CAIDA's topology of inferred inter-AS relationships. We aggregate five topologies (Oct. 14 to Feb. 15) inferred using the method presented by Luckie *et al.* [25]. We aggregate these snapshots to mitigate the impact of transient link failures on the topology used in our analysis. When inferences conflicted, we took the majority poll of inferred relationships while assigning higher weight to more recent inferences, *i.e.*, if the latest two months had the same inference, we used that inference regardless of the first three months. We use this topology to compute all paths that satisfy the Gao-Rexford (GR) model described in Section 2.

We compare the measured paths with all paths satisfying the GR model computed using CAIDA's inferred relationships. We consider two properties: (1) whether the measured path satisfies the GR model of local preference, and (2) whether the measured path has the same length as the shortest paths satisfying the GR model of local preference. Based on this we classify routing relationships as either obeying GR local preference; *i.e.*, using the neighbor with the Best Relationship type (**Best**), routing based on shortest path (**Short**), or a combination of the two.

For our active probing measurements, we consider the order in which the target AS T chooses paths. Again, we consider two properties: (1) whether the relationship between T and the next-hop on the first path is equal or better than the relationship with the next-hop on the second path, and (2) whether the first path is shorter or equal in length as the second path. We similarly label the observed decisions which obey property (1) as **Best**, and those that obey (2) as **Short**. We have limited visibility on what path the second neighbor exported to T when T chose the first path. When labeling decisions, we assume the second neighbor exported the second path to T when T chose the first path. We verified this assumption holds for the results we report.

In both cases, the sets should be treated as disjoint, with ASes that obey both Best and Short path policies appearing only in the **Best/Short** category. Observations which follow **neither** of these properties are considered inconsistent with existing models (*i.e.*, **NonBest/Long** category). There can be, however, some cases when a path suggested by CAIDA's inferences might not exist in practice. One of the reasons can be incomplete or erroneous inferences in the topologies. In addition, an AS might apply more complex filters than suggested by Gao-Rexford model when deciding which paths to advertise to neighbors (Section 4.3 discusses this in more detail).

4. HOW OFTEN DO MODELS HOLD?

We now consider how empirically observed AS paths compare with those predicted by GR model. We then investigate how often deviations can be explained by known sources of inaccuracies.

Encouragingly, we find that a majority of routing decisions (64.7%) for passively observed measurements are correctly inferred by the commonly used GR model; however, a significant fraction (34.3%) do not follow that model. Figure 1 (Simple) characterizes the observed routing decisions based on whether the path chosen is Best or Short. We find only a small number of cases (8.3%) where decisions can neither be explained by Best nor by Short path selection. In the following sections, we explore the reasons behind these decisions that differ from model-based predictions.

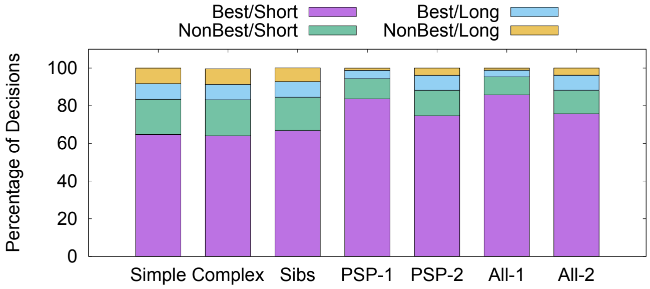


Figure 1: Breakdown of routing decisions observed by taking into account complex relationships (Complex), siblings (Sibs), prefix-specific policies (PSP-1, PSP-2) and by combining complex and siblings relationships with both criteria of prefix-specific policies (All-1, All-2).

4.1 Complex routing relationships

A well known limitation of existing routing policy models is the simplification of relationships into either customer-provider or settlement-free peering relationships. Recent work by Giotsas *et al.* addresses this limitation by augmenting relationship inferences with cases of hybrid relationships (*i.e.*, ASes whose arrangements vary based on location) and partial transit relationships (*i.e.*, ASes who will behave as providers, but only for a subset of prefixes) [15]. The hybrid relationship dataset contains pairs of ASes and the corresponding cities where relationships differ for a given AS pair. To use this dataset, we use the geolocation data from [6], which offers good coverage of infrastructure IPs such as routers. For each pair of ASes in each AS path, we geolocate corresponding IP addresses and if the geolocation data points to the same city as mentioned in hybrid relationship dataset for that AS pair, we use the hybrid relationship. Figure 1 (Complex) shows the breakdown of routing decisions observed taking into account these complex relationships. Interestingly, we find that taking these relationships into account has nearly no impact on the classification in our dataset (less than 1% change).

4.2 Sibling ASes

The mapping between AS numbers and organizations is not one-to-one [4]. Many organizations manage multiple AS numbers, either for geographic regions (e.g., Verizon with ASNs 701, 702, and 703) or due to mergers (e.g., Level 3 (AS 3356) and Global Crossing (AS 3549)).

Cai et al. [4] present a technique to map organizations to ASes by using attributes like organization IDs, email addresses and phone numbers found in `whois` information of ASes. We take a similar approach to identify AS siblings, but our approach differs in two key ways. First, we focus only on e-mail addresses in `whois` data, which previous work identified as the field with best precision and recall [4]. Second, we use DNS SOA records to identify different e-mail domains that belong to the same organization. For example, dish.com and dishaccess.tv share the dishnetwork.com authoritative domain. We also remove groups where the e-mail address is hosted by a popular e-mail provider (*e.g.*, hotmail.com), or regional Internet registry (*e.g.*, ripe.net). This results in a total of 94 sibling groups identified in our traceroute data set.

For every non GR decision that an AS makes, we check whether the AS chooses a path via a sibling. If the path is via a sibling, we mark this decision as satisfying the Best condition. Figure 1 (Sibs) shows the result of this change—3.9% more decisions are classified as Best/Short.

4.3 Prefix-specific policies

Interdomain routing is often abstracted to the level of a destination AS. However, in practice routing is based on IP prefixes which may be subject to different export policies by their originating AS (*e.g.*, forwarding prefixes hosting enterprise-class services to a more expensive provider). While Giotsas *et al.* consider partial transit [15], which is a type of prefix-specific policy, they do not explicitly consider per-prefix policies as implemented by origin ASes.

We use two criteria to identify prefix-specific policies based on correlation with BGP data obtained from Routeviews/RIPE [34, 39]. Given an origin AS (O), a neighbor N and a prefix P: **Criteria 1** do not assume the edge $N - O$ exists for prefix P unless we observe O announcing P to N in the BGP data. **Criteria 2** is similar to Criteria 1, except that we require that we observe at least one prefix announced from O to N before applying Criteria 1. The first criteria can be seen as being more aggressive whereas the second aims to ensure that our observation is actually caused by selective prefix announcement and not poor visibility.

Figure 1 (PSP-1, PSP-2) shows the breakdown of routing decisions using Criteria 1 and 2 above, respectively. We find that prefix-specific policies account for a significant fraction (10-19%) of unexpected routing decisions. Combining Criteria-1 and Criteria-2 separately with simple, complex and siblings relationships, yields 85.7% and 75.7% of decisions for Best/Short category respectively (Figure 2, All-1, All-2). One limitation of these approaches is that we only check prefix-specific policies for origin ASes. Other limitation is incomplete visibility in BGP control plane data.

Validation. In order to validate the cases of prefix-specific policies, we try to find a Looking Glass server hosted by the neighboring AS of the AS originating the prefix being examined. There were a total of 630 cases of prefix-specific policies involving 149 unique neighboring ASes. We were able to find looking glass servers in 28 of the neighboring ASes. Using these looking glass servers we manually verify 100 cases of prefix-specific policies and confirm that applying Criteria 1 was correct 78% of the time.

4.4 Active BGP Measurements

Using our active BGP measurements, we discover alternate routes. We study whether the sequence of alternate route choices match existing models and infer which step of the BGP decision

BGP DECISION	BGP FEEDS		TRACEROUTES	
Best relationship	435	(46.0%)	228	(42.4%)
Shorter path	152	(16.0%)	158	(29.4%)
Intradomain tie-breaker	155	(16.4%)	84	(15.6%)
Oldest route (magnet)	24	(2.5%)	9	(1.6%)
Violation	179	(18.9%)	58	(10.8%)
Total	**945**	(100%)	**537**	(100%)

Table 2: BGP decisions observed after we anycast a prefix previously announced from a single (magnet) location.

process led to each route. We report results for experiments performed between Feb. 25th and Apr 27th, 2015.

Alternate routes. We analyze AS routing choices when we use PEERING to discover alternate, less preferred routes. We compare the sequence of routes chosen by target ASes with CAIDA's AS-relationships database. Out of the 360 ASes we targeted, 310 (86.1%) chose routes following both Best and Shortest (as defined in Sec. 3.3); 29 (8.0%) chose routes following Best only; 18 (5.0%) following Shortest only; and 3 (0.8%) did not follow either property. We discuss the three observations that did not satisfy either property to illustrate limitations of current models.

One violation occurs for a European network E that routes via OpenPeering (AS20562)–a transit relationship identified from RPLS entries in public routing databases. After poisoning OpenPeering, E routes through (a likely peer-to-peer relationship) with AMPATH (AS20080) at AMS-IX. We list this as a violation because CAIDA identifies OpenPeering as a provider for E and AMPATH as a peer of E. Interestingly, the second route is the suffix of the first route (*i.e.*, the route through OpenPeering also reaches PEERING through AMPATH at AMS-IX), indicating the first route includes an unnecessary detour. Relationships are complex; transit and peering relationships may be preferred one over the other. Models with finer granularity for ranking neighbors of an AS may resolve these issues [27].

Another violation occurs at a US university U. The university first routes through Internet2 (AS11537) toward one of the PEERING locations in the US. After we poison Internet2, U routes through AMPATH (AS20080) toward the PEERING location in Brazil. We list this as a violation because CAIDA identifies Internet2 as a provider and AMPATH as a settlement-free peer of U. Our last observed violation is similar, where a European network first routes through Switch (AS559, identified as a provider) and then routes through NCSA (AS10764, identified as a settlement-free peer) to reach PEERING after we poison Switch. These violations indicate that identifying links used as back-up might improve our routing models.

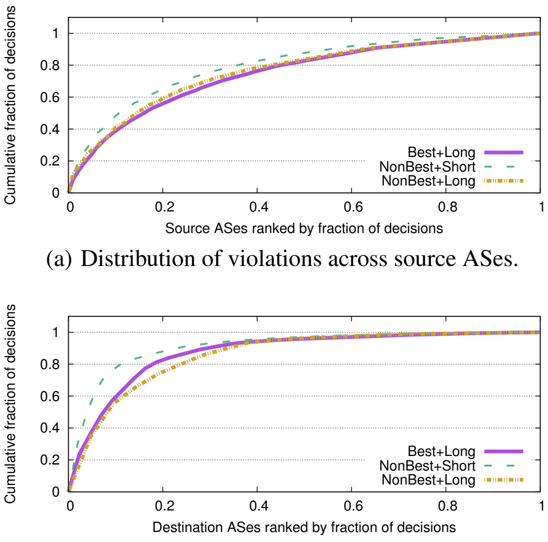
Reverse engineering BGP decisions. We now turn to our second control plane experiment, where we use anycast to explore considerations such as route age on routing decisions. Table 2 shows the root cause behind BGP routing decisions. Although most decisions are made based on relationship and path length, more than 17% of decisions are made based on intradomain tie-breakers and route age, which are not considered in and could improve current models.

Limitations. BGP poisoning does not work when BGP loop prevention is disabled or when ASes filter poisoned announcements [20,22]. Intermediate ASes between PEERING locations and target ASes may prevent us from controlling routes exported to the target AS. These factors limit our ability to identify all routes available to and neighbors of target ASes. We consider the subset of routes we observed and neighbors we identified. Moreover, our results for these experiments cover a small fraction of the Internet and are probably biased toward academic and research networks. Our control plane techniques, however, are general and could be used by other networks to cover different portions of the Internet. We believe better coverage and visibility would result in discovering more violations. To this end, we are working to extend the PEERING platform and RIPE has configured periodic measurements from a diverse set of Probes toward all PEERING prefixes.

5. SKEWNESS BY SOURCE AND DESTINATION

We now investigate which source and destination ASes account for most of the routing decisions which deviate from our model. Figure 2 (a) and (b) shows the cumulative fraction of routing decisions which violate either the Best or Short condition (*i.e.*, the AS chooses a path that is longer or more expensive than we would expect). If violations were evenly distributed across ASes, the curves would fix $y = x$; otherwise, some ASes are responsible for a disproportionately larger (or smaller) fraction of violations. We find this effect is present in both plots, but more prominently for destination ASes. We focus on the latter.

Destination ASes owned by Akamai account for 21% of violations. Of these, Cogent (AS174) is the most common source, responsible for 3.4% of these Akamai related violations. These Cogent-Akamai violations tend to occur when Cogent prefers a peer-to-peer path through a Tier-1 AS over a longer customer route towards Akamai. Netflix's AS is the destination on 17% of paths with violations. Of these, nearly a quarter (24%) are due to a stale inter-AS link in CAIDA's topology, which included a direct link between AS3549 and Netflix that no longer exists according to RIPE ASN Neighbor History [32]. For source ASes, the distribution is less skewed. Cogent and Time Warner are the top two sources, responsible for 4.1% and 2.2% of violations, respectively.



(a) Distribution of violations across source ASes.

(b) Distributions of violations across destination ASes.

Figure 2: CDF plot of the fraction of violations (x-axis) explained by source and destinations ASes (y-axis). Violations observed in our dataset are skewed significantly toward Akamai and Netflix (21% and 17% of total NonBest/Short violations respectively). The skew for source ASes is less prominent.

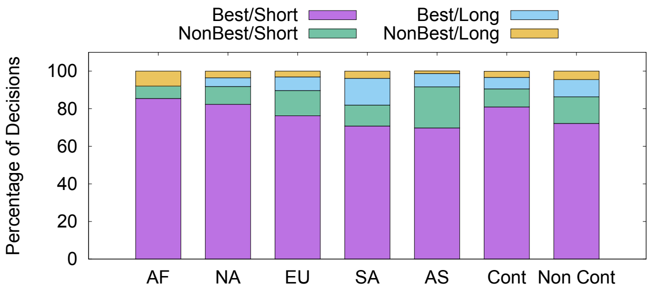


Figure 3: Breakdown of routing decisions for traceroutes that stay within continents of Africa (AF), North America (NA), Europe (EU), South America (SA), Asia (AS) and all continents combined (Cont), and for intercontinental traceroutes (Non Cont).

Continent	Non-Best/Short Decisions explained
Asia	40.1%
Africa	62.5%
Europe	64.3%
N. America	10.9%
Oceania	62.9%
S. America	66.6%

Table 3: Summary of Non-Best/Short decisions explained by ASes preferring intra-country routes.

6. IMPACT OF GEOGRAPHY

We next consider the role of geography on routing decisions. First, we isolate traceroutes that stay within a continent (Continental traceroutes), *i.e.*, all hops stay inside a given continent based on geolocating router IP addresses. Figure 3 shows the breakdown of decisions in the continental traceroutes (45% of our dataset). The fraction of decisions explained by GR for continental traceroutes is significantly greater than for intercontinental ones.

Domestic paths. Next we focused on traceroutes where we infer that the entire traceroute stayed within a single country, but there is a better *multinational Best/Short path* (in the CAIDA data), which we define to be a path with at least one AS registered (via `whois` data) in a country outside the source and destination AS's country. We find that more than 40% of non-Best/Short decisions can be explained by avoiding alternative multinational paths. One limitation of this approach is that even for the ASes that reside in multiple countries, `whois` data still points to just one country or when an AS spans across multiple regional internet registries then each RIR shows different country as the origin of that AS. Table 3 details the non-Best/Short decisions explained by ASes preferring domestic routes.

Undersea cables. Undersea cable ASes are a critical component of Internet topologies that previous works overlook [15,25]. While some cables are jointly owned by large ISPs, e.g., Pan-American Crossing, Americas-II (owned by AT&T, Sprint, and many others), we observed that others, e.g., EAC- C2C (PACNET), are operated by independent organizations using their own allocated ASNs and IP prefixes. Because these cable operators only provide point-to-point transit along the cables (i.e., they do not originate traffic and peer in locations proportional to cable landings), they resemble high-latency, high-cost IXPs and thus confuse existing AS relationship models. As such, we need techniques to identify cable ASes and correct their relationships in inferred topologies.

Violation type	Pct. of decisions explained
Non-Best & Short	3.0%
Best & Long	6.5%
Non-Best & Long	4.5%

Table 4: Fraction of decisions of each type that can be attributed to undersea cables.

We use a list of undersea cables from the TeleGeography Submarine Cable Map [38] to identify ASes for undersea cable operators. Overall, cable-ASes appear on less than 2% of paths but most of the decisions (51.2%) involving cable-ASes caused deviations from Best/Short paths. Table 4 shows fraction of each type of decision explained by undersea cable ASes.

7. CONCLUSION

In this work, we investigated how interdomain paths predicted by state-of-the-art routing models differ from empirically observed routes. We found that while a majority of paths in our dataset agree with models, more than a third do not. We explained a significant fraction of these differences due to factors such as sibling ASes, selective prefix announcements and undersea cables. Further, we investigated how the models hold up when comparing with ground-truth routing preferences revealed using PEERING announcements, and identified AS behavior that is not included in existing models. As part of future work, we are continuing to investigate cases of routing decisions that violate today's models, and we aim to incorporate our findings into new models of Internet routing.

Acknowledgments

Research in this paper was funded by NSF grants CNS-1422566, CNS-1351100, CNS-1413978; CNPq grant PU-477932; a Comcast TechFund Award; and a Google Faculty Research Award. The PEERING testbed was partially funded by NSF CNS-1406042. We gratefully acknowledge GENI for support of an early incarnation of PEERING. We also thank RIPE Atlas for help with traceroute measurements.

8. REFERENCES

[1] H. Ballani, P. Francis, and X. Zhang. A study of prefix hijacking and interception in the Internet. In *SIGCOMM*, 2007.

[2] M. Berman, J. S. Chase, L. Landweber, A. Nakao, M. Ott, D. Raychaudhuri, R. Ricci, and I. Seskar. GENI: A Federated Testbed for Innovative Network Experiments. *Computer Networks*, 61:5–23, 2014.

[3] R. Bush, O. Maennel, M. Roughan, and S. Uhlig. Internet optometry: Assessing broken glasses in internet reachability. In *ACM IMC*, 2009.

[4] X. Cai, J. Heidemann, B. Krishnamurthy, and W. Willinger. Towards an AS-to-organization map. In *ACM IMC*, 2010.

[5] M. Calder, X. F. Z. Hu, E. K.-B. J. Heidemann, and R. Govindan. Mapping the Expansion of Google's Serving Infrastructure. In *Proceedings of the ACM Internet Measurement Conference (IMC '13)*, October 2013.

[6] B. Chandrasekaran, M. Bai, M. Schoenfield, A. Berger, N. Caruso, G. Economou, S. Gilliss, B. Maggs, K. Moses, D. Duff, K. NgâĂă, E. G. Sirer, R. WeberâĂă, and B. Wong. Alidade: Ip geolocation without active probing. *Department of Computer Science, Duke University, Technical Report*, 2015.

[7] K. Chen, D. Choffnes, R. Potharaju, Y. Chen, F. Bustamante, D. Pei, and Y. Zhao. Where the sidewalk ends: Extending the internet AS graph using traceroutes from P2P users. In *CoNEXT '09*, 2009.

[8] Cisco. BGP Best Path Selection Algorithm: How the Best Path Algorithm Works. Document ID: 13753, May 2012.

[9] L. Colitti. *Internet Topology Discovery Using Active Probing*. Ph.D. thesis, University di Roma Tre, 2006.

[10] L. Gao, T. Griffin, and J. Rexford. Inherently safe backup routing with BGP. *IEEE INFOCOM*, 2001.

[11] L. Gao and J. Rexford. Stable Internet routing without global coordination. *Trans. Netw.*, 2001.

[12] P. Gill, S. Goldberg, and M. Schapira. A survey of interdomain routing policies. *ACM CCR*, 2014.

[13] P. Gill, M. Schapira, and S. Goldberg. Let the market drive deployment: A strategy for transistioning to BGP security. *SIGCOMM'11*, 2011.

[14] P. Gill, M. Schapira, and S. Goldberg. Modeling on quicksand: dealing with the scarcity of ground truth in interdomain routing data. *SIGCOMM Comput. Commun. Rev.*, 42(1):40–46, Jan. 2012.

[15] V. Giotsas, M. Luckie, B. Huffier, and K. Claffy. Inferring Complex AS Relationships. In *ACM IMC*, November 2014.

[16] S. Goldberg, M. Schapira, P. Hummon, and J. Rexford. How secure are secure interdomain routing protocols? In *SIGCOMM'10*, 2010.

[17] T. Griffin, F. B. Shepherd, and G. Wilfong. The stable paths problem and interdomain routing. *Trans. Netw.*, 2002.

[18] G. Huston. Peering and settlements - Part I. *The Internet Protocol Journal (Cisco)*, 2(1), March 1999.

[19] G. Huston. Peering and settlements - Part II. *The Internet Protocol Journal (Cisco)*, 2(2), June 1999.

[20] U. Javed, I. Cunha, D. R. Choffnes, E. Katz-Bassett, T. Anderson, and A. Krishnamurthy. Poiroot: Investigating the root cause of interdomain path changes. In *SIGCOMM*, 2013.

[21] J. Karlin, S. Forrest, and J. Rexford. Nation-state routing: Censorship, wiretapping, and BGP. *CoRR*, 2009.

[22] E. Katz-Bassett, C. Scott, D. R. Choffnes, I. Cunha, V. Valancius, N. Feamster, H. V. Madhyastha, T. Anderson, and A. Krishnamurthy. LIFEGUARD: Practical repair of persistent route failures. In *SIGCOMM*, 2012.

[23] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *SIGCOMM'10*, 2010.

[24] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang. PHAS: A prefix hijack alert system. In *Proc. USENIX Security Symposium*, 2006.

[25] M. Luckie, B. Huffaker, A. Dhamdhere, and V. Giotsas. AS relationships, customer cones, and validation. In *ACM Internet Measurement Conference*, 2013.

[26] R. Lychev, S. Goldberg, and M. Schapira. Is the juice worth the squeeze? BGP security in partial deployment. In *SIGCOMM'13*, 2013.

[27] H. Madhyastha, E. Katz-Bassett, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane Nano: Path prediction for peer-to-peer applications. In *Usenix NSDI*, 2009.

[28] R. Mazloum, M. Buob, J. Auge, B. Baynat, D. Rossi, and T. Friedman. Violation of Interdomain Routing Assumptions. In *Passive and Active Measurement Conference*, March 2014.

[29] W. Mühlbauer, A. Feldmann, O. Maennel, M. Roughan, and S. Uhlig. Building an AS-topology model that captures route diversity. In *SIGCOMM*, 2006.

[30] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. Quantifying the completeness of the observed internet AS-level structure. *UCLA Computer Science Department - Techical Report TR-080026-2008*, Sept 2008.

[31] Quantcast. `http://www.quantcast.com`.

[32] RIPE ASN Neighbor History. `https://stat.ripe.net/widget/asn-neighbours-history`.

[33] RIPE RIS raw data. `http://www.ripe.net/data-tools/stats/ris/ris-raw-data`.

[34] RIPE Network Coordination Center. RIPE Routing Information Service. http://www.ripe.net/data-tools/stats/ris/routing-information-service.

[35] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 lessons from 10 years of measuring and modeling the Internet's autonomous systems. *JSAC*, 2011.

[36] Sandvine. Fall 2012 global internet phenomena, 2012.

[37] B. Schlinker, K. Zarifis, I. Cunha, N. Feamster, and E. Katz-Bassett. PEERING: An AS for Us. In *Proc. ACM HotNets*, Los Angeles, CA, October 2014.

[38] TeleGeography Submarine Cable Map. `http://www.submarinecablemap.com/`.

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

[40] V. Valancius, N. Feamster, J. Rexford, and A. Nakao. Wide-area route control for distributed services. In *USENIX ATC*, 2010.

[41] J. Wu, Y. Zhang, Z. M. Mao, and K. Shin. Internet routing resilience to failures: Analysis and implications. In *CoNEXT*, 2007.