Are We One Hop Away from a Better Internet?

Yi-Ching Chiu (USC), Brandon Schlinker (USC), Ethan Katz-Bassett (USC), Ramesh Govindan (USC)

ABSTRACT

The Internet suffers from well-known performance and security problems. However, proposals to improve routing have seen little adoption due to the difficulty of Internet-wide deployment. We observe that, instead of trying to solve these problems in the general case, it may be possible to make substantial progress by focusing on solutions tailored to the paths between a small number of popular content providers and their clients, which carry a large share of Internet traffic.

In this paper, we demonstrate one property of these paths that may provide a foothold for deployable solutions. We find that more than 50% of end-user prefixes are directly connected to Google's network, with some other providers showing similar connectivity. Direct paths open the possibility of solutions that sidestep the headache of Internet-wide deployability, and we sketch approaches one might take to improve performance and security in this setting.

1. INTRODUCTION

Internet routing suffers from a range of problems, including slow convergence [22, 40], long-lasting outages [20], circuitous routes [38], and vulnerability to IP spoofing [5] and prefix hijacking [41]. The research and operations communities have responded with a range of proposed fixes [6, 19, 21, 27, 28]. But, despite the problems being well-known and solutions being on the table, they have seen little deployment [25, 30, 32].

One challenge for adoption is that some proposals require widespread adoption to be effective [5, 19, 25]. Such solutions are hard to deploy, since they require updates to millions of devices across tens of thousands of networks. A second challenge is that most solutions are designed to work in the general case, applicable equally to any Internet path.

We argue that, instead of solving problems for arbitrary paths, we should instead think in terms of solving problems for an arbitrary byte, query, or dollar. This solves both challenges discussed in the previous paragraph. Most traffic concentrates along a small number of routes due to a number of trends: the rise of video means that Netflix and YouTube alone account for nearly half of North American traffic [1], popular apps such as Facebook and Twitter command the attention of many users, more services are moving to shared cloud infrastructure, and a small number of mobile and broadband providers deliver Internet connectivity to most users. This skewed distribution means that an approach to improving routing can have substantial impact even if it only works over these important paths. Further, it may be possible to take advantage of properties of these paths, of the traffic along them, or of the providers using them, in order to develop tailored approaches that provide increased benefit in these scenarios even if they do not apply to arbitrary paths.

This paper focuses on one attribute of these high traffic routes: they are very short. Whereas the average path on the Internet traverses 1-2 intermediate transit ASes, our measurements show that most paths from Google go directly from Google's network into the client's network.

While previous results suggested that the Internet has been "flattening" in this way [17, 23], our results are novel in a number of ways. First, whereas previous work observed flattening in measurements sampling a small subset of the Internet, we quantify the full degree of flattening for a major content provider. Our measurements cover paths to 3.8M /24 prefixes–all of the prefixes observed to request content from a major CDN–whereas earlier work measured from only 50 [17] or 110 [23] networks. Peering links, especially of content providers like Google, are notoriously hard to uncover, with previous work projecting that traditional measurement techniques miss 90% of these links [31]. Our results support a similar conclusion to this projection: a previous study found up to 100 links per content providers across years of measurements [37] and CAIDA's analysis lists 184 Google peers [2], but we uncover links from Google to 3689 peers.

Second, we show that paths serving high volume client networks tend to be shorter than other paths. And, Google servers hosted in other networks [8] shorten paths further.

Third, beyond quantifying Google's connectivity, we provide context. ASes that Google does not peer with often have a local geographic footprint and low query volumes. Our measurements for other providers reveal that Microsoft has short peering paths similar to Google, whereas Amazon relies on Tier 1 and other providers for much of its routing.

We conclude by sketching how it might be possible to take advantage of short paths–in particular those in which the content provider peers directly with the client network–to make progress on long-standing routing problems.

- The need to work over paths that span multiple administrative boundaries caused, for example, our previous work to require complex lockstep coordination among thousands of networks [19]. How should networks coordinate when all concerned parties already have a peering relationship?

- The source and destination of traffic have direct incentive to guarantee the quality of the route between them, but intermediate networks lack visibility into end-to-end issues. With direct paths that bypass intermediate transit providers, can we design approaches that use the natural incentives of the source and destination–especially of large content providers–to actually deploy improvements?

- Some solutions designed to apply universally provide little benefit over simpler but less general techniques in likely scenarios [25]. Given the disproportionate role of a small number of providers, how much extra benefit can we achieve by tailoring our approaches to apply to these few important players?

We have not answered these questions, but we sketch problems where short paths might provide a foothold for a solution. We hope this paper will encourage the community to take advantage of the short paths of popular services to sidestep hurdles, answering these questions and others.

2. DATASETS AND DATA PROCESSING

Our measurement goal is to assess the AS path lengths between popular content providers and consumers of content. We use collections of traceroutes as well as a dataset of query volumes used to estimate the importance of paths to content.

Datasets. To assess paths from users to popular content, we use: (1) traceroutes from PlanetLab to establish a baseline of path lengths along arbitrary (not necessarily popular) routes; (2) a CDN log capturing query volumes from sets of end users; (3) traceroutes from popular cloud services to prefixes around the world; and (4) traceroutes from RIPE Atlas probes around the world to popular cloud and content providers.

Traceroutes from PlanetLab. A day of iPlane traceroutes from April 2015 [26], traceroutes from all PlanetLab sites to 154K BGP prefixes. These traceroutes represent the view of routing available from an academic testbed.

End-User Query Volumes. Aggregated and anonymized queries to a large CDN, giving (normalized) aggregate query count per /24 client prefix in one hour in 2013 across all of the CDN's globally distributed servers. The log includes queries from 3.8M client prefixes originated by 37496 ASes. The set has wide coverage, including clients in every country in the world, according to MaxMind's geolocation database .

Traceroutes from the cloud. In April and May 2015, we issued traceroutes from Google Compute Engine (GCE) [Central US region], Amazon EC2 (EC2) [Northern Virginia region], and IBM SoftLayer [Dallas DAL06 datacenter] to all 3.8M prefixes in our CDN trace and all 154K iPlane destinations. For each prefix in the CDN log, we chose a target IP address from a 2015 ISI histlist [13] to maximize the chance of a response. We issued the traceroutes using Scamper [24], which implements best practices like Paris traceroute [4].

Traceroutes from RIPE Atlas. The RIPE Atlas platform includes small hardware probes hosted in thousands of networks around the world . We issued traceroutes from Atlas probes in 1600 ASes around the world towards our cloud VMs and a small number of popular websites.

Processing traceroutes to obtain AS paths. Our measurements are IP-level traceroutes, but our analysis is over AS-level paths. Challenges exist converting to an AS path [29]. We do not innovate on this front and simply adopt widely-used practices. First, we remove any unresponsive hops, private IP addresses, and IP addresses associated with IXPs. We filter IXPs from both the IP and resulting AS paths because IXPs simply facilitate connectivity between peers; we use two lists [2, 18] to perform these conversions.

Next, we use a dataset from iPlane [26] to convert the remaining IP addresses to the ASes that originate them using BGP feeds from route collectors. We remove any Internet Exchange Points (IXPs) that appear in the path. In some cases, an IP address cannot be successfully converted to an AS; we insert an unknown AS indicator into the AS path in these scenarios. We remove one or more unknown AS indicators if the AS surrounding the unknown segments are identical, or if a single unknown AS separates two known ASes. We also merge adjacent ASes in a path if they are siblings or belong to the same organization, using existing organization lists [7], since these ASes are under shared administration.

Finally, we exclude paths that do not reach the destination AS, leaving paths to 3M /24 prefixes and the iPlane destinations. More sophisticated translation methods exist [10] but are not publicly available. Our approach adopts some of their practices, like merging siblings and removing IXPs. The remaining differences in process do not impact our qualitative conclusions, especially since our primary interest is paths that contain only one or two ASes, reducing the translation errors that may occur.

3. INTERNET PATH LENGTHS

How long are Internet paths? In this section, we demonstrate that the answer depends on the measurements used. We show that most flows from some popular web services to clients traverse at most one inter-AS link (or one *hop*), whereas traditional measurement approaches result in longer paths.

3.1 Measuring paths from the cloud

Paths from the cloud are short. As a baseline, we use our set of traceroutes from PlanetLab to iPlane destinations, as these are commonly used in academic studies. Figure 1 shows that only 3% of paths from PlanetLab are one hop to the destination, and the median path is between two and three AS hops.[1]

However, there is likely little traffic between the networks hosting PlanetLab sites (mostly universities) and most prefixes in the iPlane destination list, so these path lengths may not represent much real traffic. In fact, traffic is concentrated on a small number of links and paths from a small number of sources. For example, in 2009, 30% of traffic came from 30 ASes [23]. At a large IXP, 10% of links contribute more than 70% of traffic [35]. In contrast, many paths and links are relatively unimportant. At the same IXP, 66% of links

[1]In addition to using PlanetLab, researchers commonly use BGP route collectors to measure paths. A study of route collector archives from 2002 to 2010 found similar results to the PlanetLab traceroutes, with the average path length steadily increased from 3.65 to 3.90 [12].

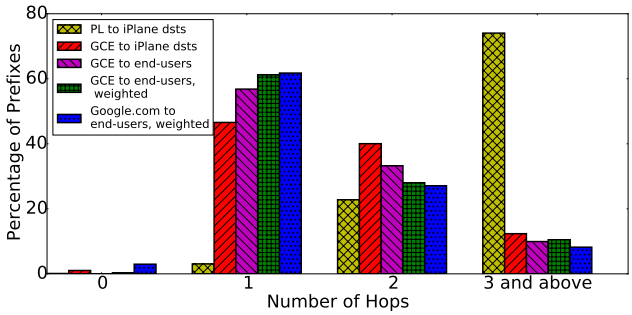

Figure 1: Paths lengths from GCE/PlanetLab to iPlane/end-user dest.

combined contributed less than 0.1% of traffic [34].

To begin answering what paths look like for one of these popular source ASes, we use our traceroutes from GCE, Google's cloud offering, to the same set of iPlane destinations. We use GCE traceroutes as a view of the routing of a major cloud provider for a number of reasons. First, traceroutes from the cloud give a much broader view than traceroutes to the cloud / content providers, since we can measure outward to all networks rather than being limited to the relatively small number where we have vantage points. Second, we are interested in the routing of high-volume services. Google itself has a number of popular services including YouTube and Search, and GCE hosts a number of third-party tenants operating popular services. The services range from latency-sensitive properties like Search to high-volume applications like YouTube, and GCE-based cloud services can benefit from the interdomain connectivity Google has established for its own services. For the majority of these services, most of the traffic flows in the outbound direction. Third, Google is at the forefront of the trends we are interesting in understanding, with open peering policies with networks around the world, their own widespread WAN [17], a cloud offering, and a rapid expansion based on ISP-hosted servers around the world [8]. Fourth, some other cloud providers that we tested filter traceroutes. §3.4 presents results from Amazon and SoftLayer, which do not filter. Finally, our previous work developed techniques that allow us to uncover the locations of Google servers and the client-to-server mapping [8], enabling some of the analysis later in this paper. Compared to PlanetLab paths, GCE paths are much shorter: 87% are at most two hop, and 47% are one hop, indicating that Google peers directly with the ASes originating the prefixes. Given the popularity of Google services in particular and cloud-based services in general, these short paths may better represent today's Internet experience.

However, even some of these paths may not reflect real traffic, as some prefixes may not host Google clients. In the rest of this section, we capture differences between GCE paths to iPlane destinations and Google paths to clients.

Paths from the cloud to end-users are even shorter. In order to understand the paths between the cloud and end-users, we use traceroutes from GCE to the 3M client prefixes in our CDN trace (§2). We assume that, since these prefixes were clients of one CDN, most of them host end-users likely to use other large web services like Google's. As seen in Figure 1, 57% of the prefixes have one hop paths from GCE, meaning their origin ASes peer directly with Google, compared to 47% of the iPlane destinations.

Prefixes with more traffic have shorter paths. The preceding analysis considers the distribution of AS hops across prefixes, but the distribution of lengths across queries/requests/flows/bytes may differ, as the per prefix volumes vary. For example, in our CDN trace, the ratio between the highest and lowest per prefix query volume is 8.7M:1. To approximate the number of AS hops experienced by queries, the *GCE to end-users, weighted* line in Figure 1 weights each of the 3M prefixes by its query volume in our CDN trace (§2), with over 60% of the queries coming from prefixes with a one hop path. The dataset has limitations: it is only an hour, so it suffers time-of-day distortions, and prefix weights are representative of the CDN's client distribution but not necessarily Google's client distribution. The dataset suffices for our purposes: precise ratios are not as important as the trends of how paths with no/low traffic differ from paths with high traffic, and a prefix that originates many queries in this dataset is more likely to host users generating many queries for other web services. While our quantitative results would differ with a trace from a different provider, we believe that qualitative differences between high and low volume paths would hold.

3.2 Estimating paths to a popular service

The previous results measured the length of paths from Google's GCE cloud service towards end-user prefixes. However, these paths may not be the same as the paths from large web properties such as Google Search and YouTube for at least two reasons. First, Google and some other providers deploy front-end servers inside some end-user ASes [8], which we refer to as *off-net* servers. So, some client connections terminate at off-nets hosted in other ASes than where our GCE traceroutes originate. Second, it is possible that Google uses different paths for its own web services than it uses for GCE tenants. In this section, we first describe how we estimate the number of AS hops for clients to `google.com`, considering both those factors. We then validate our approach. Finally, we use our approach to estimate the number of AS hops from clients to `google.com` and show that some of the paths are shorter than our GCE measurements above.

Estimating Number of AS Hops to Google Search. First, we use EDNS0 client-subnet queries to resolve `google.com` for each /24 end-user prefix, as in our previous work [8]. Each query returns a set of server IP addresses for that end-user prefix to use. Next, we translate the server addresses into ASes as described in §2. We discard any end-user prefix that maps to servers in multiple ASes, leaving a set of prefixes directed to servers in Google's AS and a set of prefixes directed to servers in other ASes.

For end-user prefixes directed towards Google's AS, we estimate the number of AS hops to `google.com` as equal to the number of AS hops from GCE to the end-user prefix, un-

Table 1: Estimated vs. measured path lengths from Atlas to `google.com`

Type	Count	no error	error ≤ 1 hop
paths to on-nets	1,087	83.56%	98.25%
paths to off-nets	224	90.18%	96.97%
paths of length ≤ 2	655	88.24%	98.62%

der the assumption, which we will later validate, that Google uses similar paths for its cloud tenants and its own services. For all other traces, we build a graph of customer/provider links in CAIDA's AS relationship dataset [2] and estimate the number of AS hops as the length of the shortest path between the end-user AS and the off-net server's AS.[2] Since off-net front-ends generally serve only clients in their customer cone [8] and public views such as CAIDA's should include nearly all customer/provider links that define these customer cones [31], we expect these paths to usually be accurate.

Validating Estimated AS Hops: To validate our methodology for estimating the number of AS hops to `google.com`, we issued traceroutes from 1,311 RIPE Atlas probes to `google.com` and converted them to AS paths. We also determined the AS hosting the Atlas probe and estimated the number of AS hops from it to `google.com` as described above.[3]

For the 224 ground-truth traces directed to off-nets, we calculate the difference between the estimated and measured number of AS hops. For the remaining 1,087 traces that were directed to front-ends within Google's network, we may have traceroutes from GCE to multiple prefixes in the Atlas probe's AS. If their lengths differed, we calculate the difference between the Atlas-measured AS hops and whichever GCE-measured path had the closest number of AS hops.

§3.2 shows the result of our validation: overall, 85% of our estimates have the same number of AS hops as the measured paths, and 88% in cases where the number of hops is one (front-end AS peers with client AS). We conclude that our methodology is accurate enough to estimate the number of AS hops for all clients to `google.com`, especially for the short paths we are most interested in.

Off-net front-ends shorten some paths even more. Applying our estimation technique to the full set of end-user prefixes, we arrive at the estimated AS hop distribution shown in the *Google.com to end-users, weighted* line in Figure 1. The estimated paths between `google.com` and end-user prefixes are shorter overall than the traces from GCE, with 67% of queries coming from ASes that either peer with Google, use off-nets hosted in their providers, or themselves host off-nets. For clients served by off-nets, the front-end to back-end portion of their connections also crosses domains, starting in the hosting AS and ending in a Google datacenter. The connection from the client to front-end likely plays the largest role in client-perceived performance, and Google has more control over the front-end to back-end connection [14], but it is worth briefly considering that leg of the split connection. We issued

[2]If the end-user AS and off-net AS are the same, the length is zero.

[3]We are unable to determine the source IP address for some Atlas probes and thus make estimations at the AS level

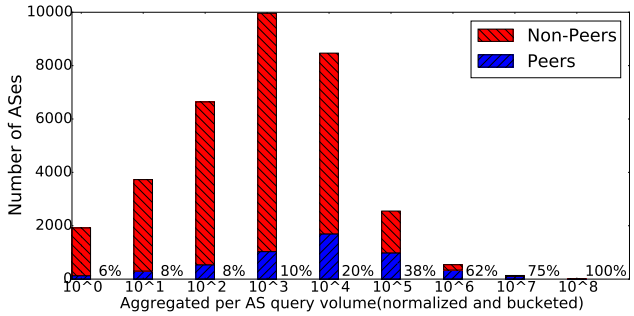

Figure 2: How many (and what fraction) of ASes Google peers with by AS size. AS size is the number of queries that flow through it, given paths from GCE to end-user prefixes and per prefix query volumes in our CDN trace. Volumes are normalized and bucketed by powers of 10.

traceroutes from GCE to the full set of Google off-nets [8]. Google has a direct connection to the hosting AS for 62% of off-nets, and there was only a single intermediate AS for an additional 34%.

3.3 Google's Peers (and Non-Peers)

In our traceroutes from GCE, we observed Google peering with 3689 ASes (after merging siblings).[4] Since a primary reason to peer is to reduce transit costs, we first investigate the projected query volume of ASes that do and do not peer with Google. We form a flow graph by combining the end-user query volumes from our CDN trace with the AS paths defined by our GCE traceroutes. So, for example, the total volume for an AS will have both the queries from that AS's prefixes and from its customer's prefixes if traceroutes to the customer went via the AS. We group the ASes into buckets based on this aggregated query volume, and Figure 2 shows the number of ASes within each bucket that do / do not peer with Google in our traceroutes. As expected, Google peers with a larger fraction of higher volume ASes. There are still high volume ASes that do not peer with Google. However, most ASes that do not peer are small in terms of traffic volume and, up to the limitations of public geolocation information, geographic footprint. We used MaxMind to geolocate the prefixes that Google reaches via a single intermediate transit provider, then grouped those prefixes by origin AS. Of 20946 such ASes, 74% have all their prefixes located within a 50 mile diameter.[5] However, collectively these ASes account for only 4% of the overall query volume.

Connectivity also varies by region. For example, overall, 10% of the queries in our CDN log come from end users in China, 25% from the US, and 20% from Asia-Pacific excluding China. However, China has longer paths and less direct peering, so 27% of the 2 hop paths come from China, and only 15% from the US and 10% from Asia-Pacific.

3.4 Paths to Other Popular Content

In this section, we compare measurements of Google and

[4]For the interested reader, Google publishes its peering policy and facilities list on a website and in PeeringDB .

[5]Geolocation errors may distort this result, although databases tend to be more accurate for end-user prefixes like the ones in question.

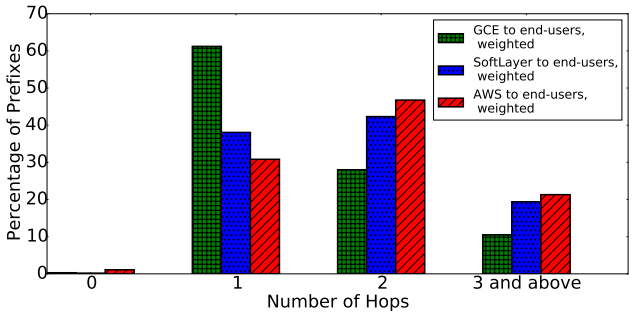

Figure 3: Paths lengths from different cloud platforms to end-users.

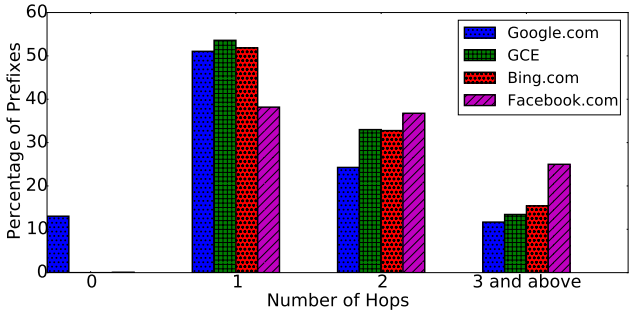

Figure 4: Path lengths from RIPE Atlas nodes to content and cloud

other providers. First, in Figure 3, we compare the number of AS hops from GCE to the end-user prefixes (weighted by query volume as in the top lines from Figure 1) to the number of AS hops to the same targets from two other cloud providers. SoftLayer and AWS have similar path distributions to each other but differ from GCE. While both SoftLayer and AWS have a substantial number of one hop paths, both are under 40%, compared to more than 60% for GCE. Still, the vast majority of both SoftLayer and AWS paths have two hops or less. Our measurements and related datasets suggest that these three cloud providers employ different strategies from each other: Google peers widely, with 3689 next hop ASes in our traceroutes,[6] and only has 5 providers in CAIDA data [2], using routes through those providers to reach end users responsible for 10% of the queries in our CDN trace; Amazon only has 329 next hop ASes, but has 20 providers that it uses for routes to 50% of the end user queries; and SoftLayer is a middle ground, with 1184 next hops and 11 providers it uses to reach end users with 47% of the queries.

We anticipate that some other large content providers are building networks similar to Google's to reduce transit costs and improve quality of service for end-users. Since we cannot issue traceroutes from within these providers' networks towards end-users, we use traceroutes from RIPE Atlas vantage points towards the providers. We execute traceroutes from a set of Atlas probes towards `facebook.com` and Microsoft's `bing.com`. We calibrate these results with our earlier ones by comparing to traceroutes from the Atlas probes towards `google.com` and our GCE instance.

Figure 4 shows the number of AS hops to each destination. The AS hop distribution to `bing.com` is nearly identical to the AS hop distribution to GCE. Paths to `bing.com` are longer than paths to `google.com`, likely because Microsoft does not have an extensive set of off-net servers like Google's. Some paths to `facebook.com` are longer, with just under 40% of probes having 1 AS hop.

Summary. *Path lengths for popular services tend to be much shorter than random Internet paths. For instance, while only 3% of PlanetLab paths to iPlane destinations are one hop, we estimate that 67% of queries to* `google.com` *go directly from the client AS to Google.*

4. CAN SHORT PATHS BE BETTER PATHS?

Our measurements suggest that much of the Internet's popular content flows across at most one interdomain link on its path to clients. In this section, we argue that these direct connections may represent an avenue to making progress on long-standing Internet routing problems. Within the confines of this short paper, we cannot develop complete solutions. Instead, we sketch where and why progress may be possible, starting with general arguments about why short paths may help, and then continuing with particular problems where short paths may yield deployable solutions. We hope this paper serves as a spark for future work in this area.

4.1 Short paths sidestep existing hurdles

Paths to popular content will become short. Competitive pressures and the need to ensure low latency access to popular content will continue to accelerate this trend. Services are moving to well-connected clouds; providers are building out serving infrastructure [8, 15]; peering is on the rise [9, 34]; even the long-tail of websites can be proxied through Google servers. The rise of video [1] and interactive applications suggests that providers will continue to seek peering and distributed infrastructure to reduce costs and latency. Because traffic is concentrating along short paths, solutions tailored for this setting can have impact, even if they do not work for or do not achieve deployment along arbitrary Internet paths.

One-hop paths only involve invested parties. The performance of web traffic depends on the intra- and interdomain routing decisions of every AS on the path. The source and destination have incentives to improve performance, as it impacts their quality of experience and revenue. Transit ASes in the middle are not as directly invested in the quality of the route. One-hop paths bypass transit, and the only ASes are senders and receivers with motivation to improve routing.

Internet protocols support communication between neighbors. An AS can use MEDs, selective advertisements, and BGP communities to express policy that may impact the routing of neighbors. ASes are willing to work together [38] using these mechanisms. However, the mechanisms generally only allow communication to neighbors: MEDs and communities usually do not propagate past one hop [33], and selective advertising only controls which neighbors receive a route. This limitation leaves ASes with almost no ability to affect the routing past their immediate neighbors,[7] but

[6] All numbers after merging siblings and organizations.

[7] Some ASes offer communities to influence route export.

one-hop paths only consist of immediate neighbors.

4.2 Short paths can simplify many problems

Joint traffic engineering. BGP does not support the negotiation of routing and traffic engineering between autonomous systems. Instead, network operators *hint* via MEDs and prepending, to indicate to neighbor ASes their preferences for incoming traffic. The coarse granularity of these hints and the lack of mechanisms to mutually optimize across AS boundaries result in paths with inflated latencies [38].

Prior work proposed protocols to jointly optimize routing between neighboring ASes [28]. Yet such protocols become more complex when they must be designed to optimize paths that traverse remote facilities and intermediary ASes [27], to the point that it is unclear what fairness and performance properties they guarantee. In comparison, one-hop paths between provider and end-user ASes reduce the need for complicated solutions, enabling direct negotiation between the parties that benefit the most. Since the AS path is direct and does not involve the rest of the Internet, it may be possible to use channels or protocols outside or alongside BGP, without requiring widespread adoption of changes.

Preventing spoofed traffic. Major barriers exist to deploying effective spoofing prevention. First, filters are only easy to deploy correctly near the edge of the Internet [5]. Second, existing approaches do not protect the AS deploying a filter, instead protecting that AS from originating attacks on others. So, ASes lack strong incentives to deploy spoofing filters [5].

The short paths on today's Internet create a setting where it may be possible to protect against spoofing attacks for large swaths of the Internet by sidestepping the existing barriers. A cloud provider like Google that connects directly to most origins should know valid source addresses for traffic over any particular peering and be able to filter spoofed traffic, perhaps using strict uRPF filter. The direct connections address the first barrier by removing the routing complexity that complicates filter configuration, essentially removing the core of the Internet from the path entirely. The cloud provider is the destination,[8] removing the second barrier as it can protect itself by filtering spoofed traffic over many ingress links. While these mechanisms do not prevent all attacks,[9] they reduce the attack surface and may be part of a broader solution.

Limiting prefix hijacks. Prefix origins can be authenticated with the RPKI, now being adopted, but it does not enable authentication of the non-origin ASes along a path [25]. So, a provider having direct paths does not on its own prevent other ASes from hijacking the provider's prefixes via longer paths. While RPKI plus direct paths are not a complete solution by themselves, we view them as a potential building block towards more secure routing. If an AS has authenticated its prefix announcements–especially an important content provider or set of end users–it seems reasonable for direct peers to configure preferences to prefer one-hop, RPKI-validated announcements over competing advertisements.

Speeding route convergence. BGP can experience delayed convergence [22], inspiring general clean-slate alternatives such as HLP [39] and simpler alternatives with restricted policies that have better convergence properties. Our findings on the flattening of the path distribution may make the latter class of solutions appealing. Specifically, it may suffice to deploy restricted policies based on BGP next-hop alone [36] for one-hop neighbors. In this as well, the incentive structure is just right: delayed route failovers can disrupt popular video content, so the content provider wants to ensure fast failover to improve the user's quality of experience.

Avoiding outages. The Internet is susceptible to long-lasting partial outages in transit ASes [20]. The transit AS lacks visibility into end-to-end connections, so may not detect a problem, and the source and destination lack visibility into or control over transit ASes, making it difficult to even discern the location of the problem [21]. With a direct path, an AS has much better visibility and control over its own routing to determine and fix a local problem, or it can know the other party–also invested in the connection–is to blame. Proposals exist to enable coordination between providers and end-user networks [16], and such designs could enable reactive content delivery that adapts to failures and changes in capacity.

5. RELATED WORK

Previous work observed a trend towards flattening and the centrality of content using active measurements [17], passive monitoring [3,23], and modeling [11]. Our work extends this observation by measuring to and from a much larger number of networks. Earlier work showed advantages to allowing direct negotiation between neighbors [28] and CDNs and access ISPs [16], similar to approaches we imagine over direct paths. Work showing the benefits to BGP policy based only on the next hop [36] helps demonstrate the potential of such approaches.

6. CONCLUSIONS

As large content and cloud providers have been building out content distribution infrastructure and engaging in direct peering, a majority of clients are one AS hop away from important content. This trend towards one-hop paths for important content will likely accelerate, driven by competitive pressures, and by the need to reduce latency for improved user experience. This suggests that, in a departure from the current focus on general solutions, interdomain routing and traffic management techniques should focus on optimizing for the common case of one-hop paths, a regime where simpler, deployable solutions may exist.

7. REFERENCES

[1] Sandvine global internet phenomena report, 2014.

[2] The caida as relationships dataset. `http://www.caida.org/data/as-relationships/`, cited February 2015.

[8] Most cloud and content providers are stub networks.

[9] An attacker can still spoof as a cloud provider in a reflection attack.

[3] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig. Web content cartography. IMC '11.
[4] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding traceroute anomalies with paris traceroute. IMC '06.
[5] R. Beverly, A. Berger, Y. Hyun, and k. claffy. Understanding the efficacy of deployed internet source address validation filtering. IMC '09.
[6] K. Butler, T. R. Farley, P. McDaniel, and J. Rexford. A survey of bgp security issues and solutions. *Proceedings of the IEEE*, 2010.
[7] X. Cai, J. Heidemann, B. Krishnamurthy, and W. Willinger. An organization-level view of the Internet and its implications (extended). Technical report, USC/ISI TR, June 2012.
[8] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the expansion of google's serving infrastructure. IMC '13.
[9] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There is more to ixps than meets the eye. *ACM SIGCOMM CCR*.
[10] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao. Where the sidewalk ends: Extending the internet as graph using traceroutes from p2p users. CoNEXT '09.
[11] A. Dhamdhere and C. Dovrolis. The internet is flat: Modeling the transition from a transit hierarchy to a peering mesh. Co-NEXT '10.
[12] B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet topology over time. *arXiv preprint*, 2012.
[13] X. Fan and J. Heidemann. Selecting representative ip addresses for internet topology studies. IMC '10.
[14] T. Flach, N. Dukkipati, A. Terzis, B. Raghavan, N. Cardwell, Y. Cheng, A. Jain, S. Hao, E. Katz-Bassett, and R. Govindan. Reducing web latency: The virtue of gentle aggression. SIGCOMM '13.
[15] A. Flavel, P. Mani, D. Maltz, N. Holt, J. Liu, Y. Chen, and O. Surmachev. Fastroute: A scalable load-aware anycast routing architecture for modern cdns. NSDI '15.
[16] B. Frank, I. Poese, Y. Lin, G. Smaragdakis, A. Feldmann, B. Maggs, J. Rake, S. Uhlig, and R. Weber. Pushing cdn-isp collaboration to the limit. *ACM SIGCOMM CCR*.
[17] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse? PAM'08.
[18] V. Giotsas, M. Luckie, B. Huffaker, and k. claffy. Inferring complex as relationships. IMC '14.
[19] J. P. John, E. Katz-Bassett, A. Krishnamurthy, T. Anderson, and A. Venkataramani. Consensus routing: The internet as a distributed system. NSDI'08.
[20] E. Katz-Bassett, H. V. Madhyastha, J. P. John, A. Krishnamurthy, D. Wetherall, and T. Anderson. Studying black holes in the internet with hubble. NSDI'08.
[21] 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. SIGCOMM '12.
[22] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed internet routing convergence.
[23] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. SIGCOMM '10.
[24] M. Luckie. Scamper: A scalable and extensible packet prober for active measurement of the internet. IMC '10.
[25] R. Lychev, S. Goldberg, and M. Schapira. Bgp security in partial deployment: Is the juice worth the squeeze? SIGCOMM '13.
[26] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iplane: An information plane for distributed services. OSDI '06.
[27] R. Mahajan, D. Wetherall, and T. Anderson. Mutually controlled routing with independent isps. NSDI'07.
[28] R. Mahajan, D. Wetherall, and T. Anderson. Negotiation-based routing between neighboring isps. NSDI'05.
[29] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz. Towards an accurate as-level traceroute tool. SIGCOMM '03.
[30] J. Mirkovic and E. Kissel. Comparative evaluation of spoofing defenses. *Dependable and Secure Computing, IEEE Transactions on*, March 2011.
[31] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (in)completeness of the observed internet as-level structure. *IEEE/ACM Trans. Netw.*
[32] S. Peter, U. Javed, Q. Zhang, D. Woos, T. Anderson, and A. Krishnamurthy. One tunnel is (often) enough. SIGCOMM '14.
[33] B. Quoitin and O. Bonaventure. A survey of the utilization of the bgp community attribute. Internet-Draft draft-quoitin-bgp-comm-survey-00, February 2002.
[34] P. Richter, G. Smaragdakis, A. Feldmann, N. Chatzis, J. Boettger, and W. Willinger. Peering at peerings: On the role of ixp route servers. IMC '14.
[35] M. A. Sanchez, F. E. Bustamante, B. Krishnamurthy, W. Willinger, G. Smaragdakis, and J. Erman. Inter-domain traffic estimation for the outsider. IMC '14.
[36] M. Schapira, Y. Zhu, and J. Rexford. Putting bgp on the right path: A case for next-hop routing. Hotnets-IX, 2010.
[37] Y. Shavitt and U. Weinsberg. Topological trends of internet content providers. SIMPLEX '12.
[38] N. Spring, R. Mahajan, and T. Anderson. The causes of path inflation. SIGCOMM '03.
[39] L. Subramanian, M. Caesar, C. T. Ee, M. Handley, M. Mao, S. Shenker, and I. Stoica. Hlp: A next generation inter-domain routing protocol. SIGCOMM '05.
[40] F. Wang, Z. M. Mao, J. Wang, L. Gao, and R. Bush. A measurement study on the impact of routing events on end-to-end internet path performance.
[41] Z. Zhang, Y. Zhang, Y. C. Hu, Z. M. Mao, and R. Bush. Ispy: detecting ip prefix hijacking on my own. In *SIGCOMM 08'*.