Are We One Hop Away from a Better Internet?

Yi-Ching Chiu*, Brandon Schlinker*, Abhishek Balaji Radhakrishnan,

Ethan Katz-Bassett, Ramesh Govindan

Department of Computer Science, University of Southern California

ABSTRACT

The Internet suffers from well-known performance, reliability, and security problems. However, proposed improvements have seen little adoption due to the difficulties 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 popular content providers and their clients, which carry a large share of Internet traffic.

In this paper, we identify one property of these paths that may provide a foothold for deployable solutions: they are often very short. Our measurements show that Google connects directly to networks hosting more than 60% of end-user prefixes, and that other large content providers have similar connectivity. These 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.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Network topology; C.2.5 [**Local and Wide-Area Networks**]: Internet

Keywords

Measurements; Internet topology

1. INTRODUCTION

Internet routing suffers from a range of problems, including slow convergence [25, 43], long-lasting outages [23], circuitous routes [41], and vulnerability to IP spoofing [6] and prefix hijacking [44]. The research and operations communities have responded with a range of proposed fixes [7, 22, 24, 30, 31]. However, proposed solutions to these well-known problems have seen little adoption [28, 33, 35].

One challenge is that some proposals require widespread adoption to be effective [6, 22, 28]. 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 the goal is often an approach that works in the general case, applicable equally to any Internet path, and it may be difficult to design such general solutions.

We argue that, instead of solving problems for arbitrary paths, we can think in terms of solving problems for an arbitrary byte, query, or dollar, thereby putting more focus on paths that carry a higher volume of traffic. Most traffic concentrates along a small number of routes due to a number of trends: the rise of Internet video had led to Netflix and YouTube alone accounting for nearly half of North American traffic [2], more services are moving to shared cloud infrastructure, and a small number of mobile and broadband providers deliver Internet connectivity to end-users. This skewed distribution means that an approach to improving routing can have substantial impact even if it only works well 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.

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

While previous results suggested that the Internet has been "flattening" in this manner [20, 26], 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 [20] or 110 [26] 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 [34]. Our results support a similar conclusion to this projection: Whereas a previous study found up to 100 links per content provider across years of measurements [40] and CAIDA's analysis lists 184 Google peers [3], our analysis uncovers links from Google to over 5700 peers.

Second, we show that, from the same content provider, popular paths serving high volume client networks tend to be shorter than paths to other networks. Some content providers even host servers in other networks [9], which in effect shortens paths further.

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

*These authors contributed equally to this work.

IMC'15, October 28–30, 2015, Tokyo, Japan.

DOI: http://dx.doi.org/10.1145/2815675.2815719.

We conclude by asking whether 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. Is it easier to make progress in this setting that, while limited, holds for much of our web activity?

- The need to work over paths that span multiple administrative boundaries caused, for example, our previous route reliability work to require complex lockstep coordination among thousands of networks [22]. Is coordination simplified 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 deploy improvements?
- Some solutions designed to apply universally provide little benefit over simpler but less general techniques in likely scenarios [28]. Given the disproportionate role of a small number of providers, can we achieve extra benefit 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 and improve Internet routing.

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 to estimate the importance of different paths.

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 end users around the world; (3) traceroutes from popular cloud service providers 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 [29], which contains 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 [1].

While the exact per prefix volumes would vary across provider and time, we expect that the trends shown by our results would remain similar. To demonstrate that our CDN log has reasonable query distributions, we compare it with a similar Akamai log from 2014 (Fig. 21 in [16]). The total number of /24 prefixes requesting content in the Akamai log is 3.76M, similar to our log's 3.8M prefixes. If V_n^C and V_n^A are the percentage of queries from the top n prefixes in our CDN dataset and in Akamai's dataset, respectively, then $|V_n^C - V_n^A| < 6\%$ across all n. The datasets are particularly similar when it comes to the contribution of the most active client prefixes: $|V_n^C - V_n^A| < 2\%$ for $n \leq 100,000$, which accounts for ≈31% of the total query volume.

Traceroutes from the cloud. In March and August/September 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 hitlist [15] to maximize the chance of a response. We issued the traceroutes using Scamper [27], which implements best practices like Paris traceroute [5].

Traceroutes from RIPE Atlas. The RIPE Atlas platform includes small hardware probes hosted in thousands of networks around the world. In April 2015, we issued traceroutes from Atlas probes in approximately 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 in converting IP-level paths to AS paths [32]; 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 Internet Exchange Points (IXPs).[1]

Next, we use a dataset from iPlane [29] to convert the remaining IP addresses to the ASes that originate them, and we remove any ASes that correspond to IXPs. If the iPlane data does not include an AS mapping for an IP address, we insert an unknown AS indicator into the AS path. We remove one or more unknown AS indicators if the ASes on both sides of the unknown segment are the same, or if a single unknown hop separates two known ASes. After we apply these heuristics, we discard paths that still contains unknown segments. We then merge adjacent ASes in a path if they are siblings or belong to the same organization, using existing organization lists [8], since these ASes are under shared administration.[2]

Finally, we exclude paths that do not reach the destination AS. For our traceroutes from the cloud, we are left with paths to 3M of the 3.8M /24 prefixes.

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 datasets 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 traceroutes from academic testbeds are commonly used in academic studies.

[1]We filter IXPs because they simply facilitate connectivity between peers. We use two CAIDA supplementary lists [3, 21] to identify ASes and IPs associated with IXPs.

[2]We compared the AS paths inferred by our approach with those inferred by a state-of-the-art approach designed to exclude paths with unclear AS translations [12], generating the results in our paper using both approaches. The minor differences in the output of the two approaches do not impact our results meaningfully, and so we only present results from our approach.

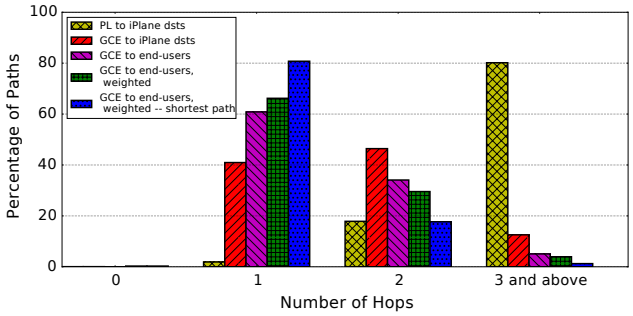

Figure 1: Paths lengths from GCE/PlanetLab to iPlane/end-user dest.

Figure 1 shows that only 2% of paths from PlanetLab are one hop to the destination, and the median path is between two and three AS hops.[3] However, there is likely little traffic between the networks hosting PlanetLab sites (mostly universities) and most prefixes in the iPlane list, so these longer paths may not carry much traffic.

Instead, 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 [26]. At a large IXP, 10% of links contribute more than 70% of traffic [38]. In contrast, many paths and links are relatively unimportant. At the same IXP, 66% of links combined contributed less than 0.1% of traffic [37].

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* cloud and 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, ranging from latency-sensitive properties like Search to high-volume applications like YouTube. GCE also hosts a number of third-party tenants operating popular services which 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 interested in understanding, maintaining open peering policies around the world, a widespread WAN [20], a cloud offering, and ISP-hosted front end servers [9]. Fourth, some other cloud providers that we tested filter traceroutes (§3.4 discusses measurements from Amazon and SoftLayer, which also 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 [9], enabling some of the analysis later in this paper.

Compared to PlanetLab paths towards iPlane destinations, GCE paths are much shorter: 87% are at most two hop, and 41% 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 iPlane prefixes may not host Google clients. In the rest of this section and §3.3, we capture differences in Google's and GCE's paths toward iPlane destinations and end-users.

[3]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 number of hops *increasing* from 2.65 to 2.90 [14].

Paths from the cloud to end-users are even shorter. In order to understand the paths between the cloud and end-users, we analyze 3M traceroutes from GCE to client prefixes in our CDN trace (§2). We assume that, since these prefixes contain 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, 61% of the prefixes have one hop paths from GCE, meaning their origin ASes peer directly with Google, compared to 41% 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 across queries/requests/flows/bytes may differ, as 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 66% of the queries coming from prefixes with a one hop path.

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. The dataset has limitations: the trace is only one hour, so suffers time-of-day distortions, and prefix weights are representative of the CDN's client distribution but not necessarily Google's client distribution. However, 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 services.

Path lengths to a single AS can vary. We observed traceroutes traversing paths of different lengths to reach different prefixes within the same destination AS. Possible explanations for this include: (1) traffic engineering by Google, the destination AS, or a party in between; and (2) split ASes, which do not announce their entire network at every peering, often due to lack of a backbone or a capacity constraint. Of 17,905 ASes that had multiple traceroutes in our dataset, 4876 ASes had paths with different lengths. Those 4876 ASes contribute 72% of the query volume in our CDN trace, with most of the queries coming from prefixes that have the shortest paths for the ASes. The *GCE to end-users weighted, shortest path* bars in Figure 1 show how long paths would be if all traffic took the shortest observed path to its destination AS. With this hypothetical routing, 80% of queries traverse only one hop.

Path lengths vary regionally. Peering connectivity can also vary 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.

3.2 Google's Peers (and Non-Peers)

In our traceroutes from GCE, we observed Google peering with 5083 ASes (after merging siblings).[4,5] 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.

[4]For the interested reader, Google publishes its peering policy and facilities list at `http://peering.google.com` and in PeeringDB.
[5]Some of these peers may be using remote peering [10].

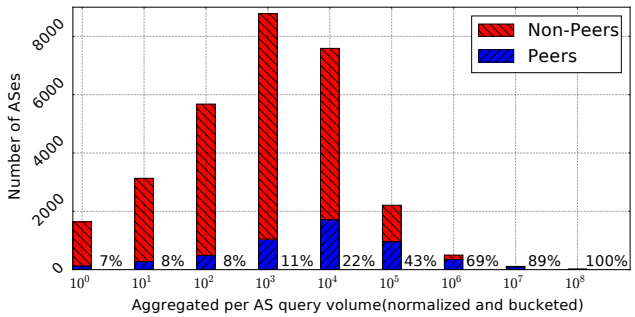

Figure 2: How many (and what fraction) of ASes Google peers with by AS size. AS size is the number of queries that flow through it, given paths from GCE to end-user prefixes and per prefix query volumes in our CDN trace. Volumes are normalized and bucketed by powers of 10.

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. And while there are still high volume ASes that do not peer with Google, 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 20,946 such ASes, 74% have all their prefixes located within a 50 mile diameter.[6] However, collectively these ASes account for only 4% of the overall query volume.

Peering is increasing over time. We evaluated how Google's visible peering connectivity changed over time by comparing our March 2015 traces with an additional measurement conducted in August 2015. In August, we observed approximately 700 more peerings than the 5083 we measured in March. While some of these peerings may have been recently established, others may have been previously hidden from our vantage point, possibly due to traffic engineering. These results suggest that a longitudinal study of cloud connectivity may provide new insights.

3.3 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 [9], which we refer to as *off-net* servers. As a result, 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 path length from end-users to `google.com`, considering both of these factors. We then validate our approach. Finally, we use our approach to estimate the number of AS hops from end-users to `google.com` and show that some of the paths are shorter than our GCE measurements above.

Estimating 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 [9]. 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.

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

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

Type	Count	no error	error ≤ 1 hop
all paths	1,409	81.26%	97.16%
paths to on-nets	1,120	80.89%	98.21%
paths to off-nets	289	82.70%	93.08%
paths w/ ≤ 1 hop	925	86.05%	97.62%

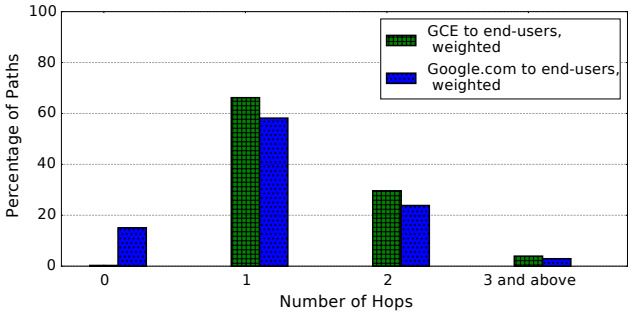

Figure 3: Paths lengths from Google.com and GCE to end-users

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, under 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 connectivity in CAIDA's AS relationship dataset [3] 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.[7] Since off-net front-ends generally serve only clients in their customer cone [9] and public views such as CAIDA's should include nearly all customer/provider links that define these customer cones [34], 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 used traceroutes from 1409 RIPE Atlas probes[8] 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.[9]

For the 289 ground-truth traces directed to off-nets, we calculate the difference between the estimated and measured number of AS hops. For the remaining 1120 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 the GCE-measured path with the closest number of AS hops.

Table 1 shows the result of our validation: overall, 81% of our estimates have the same number of AS hops as the measured paths, and 85% 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 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 3.

The estimated paths between `google.com` and end-user prefixes are shorter overall than the traces from GCE, with 73% 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 portions of their connections also

[7]If the end-user AS and off-net AS are the same, the length is zero.
[8]The rest of the 1600 traceroutes failed.
[9]We are unable to determine the source IP address for some Atlas probes and thus make estimations at the AS level.

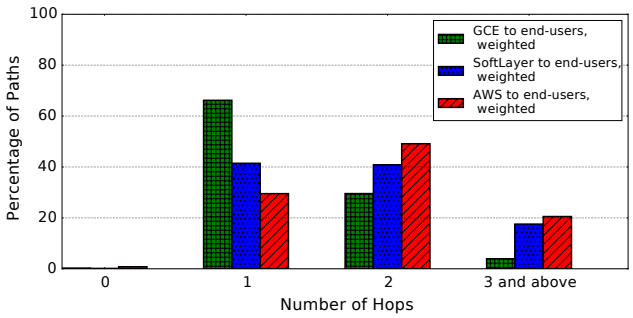

Figure 4: Paths lengths from different cloud platforms to end-users.

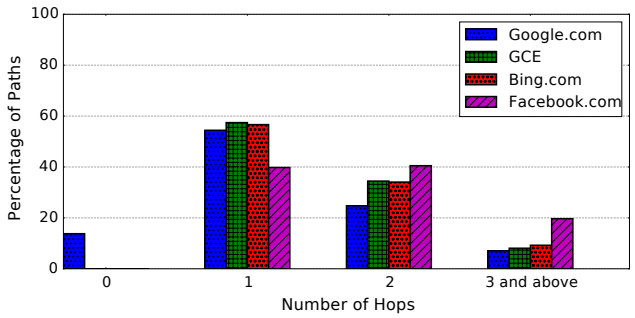

Figure 5: Path lengths from RIPE Atlas nodes to content and cloud[11]

cross 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, since Google has greater control of, and can apply optimizations to, the connection between the front-end and back-end [17]. Still, we evaluated that leg of the split connection by issuing traceroutes from GCE to the full set of Google off-nets [9]. Our measurements show that 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%.

3.4 Paths to Other Popular Content

In this section, we compare measurements of Google and other providers. First, in Figure 4, we compare the number of AS hops (weighted by query volume) from GCE to the end-user prefixes to the number of AS hops to the same targets from two other cloud providers. While SoftLayer and AWS each have a substantial number of one hop paths, both are under 42%, compared to well over 60% for GCE. Still, the vast majority of 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 5083 next hop ASes in our traceroutes, and only has 5 providers in CAIDA data [3], using routes through those providers to reach end users responsible for 10% of the queries in our CDN trace; Amazon only has 756 next hop ASes, but uses 20 providers for routes to 50% of the end user queries; and SoftLayer is a middle ground, with 1986 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,[10] 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 5 shows the number of AS hops to each destination.[11] 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. Facebook trails the pack, with just under 40% of paths to `facebook.com` having 1 AS hop.

Summary. *Path lengths for popular services tend to be much shorter than random Internet paths. For instance, while only 2% of PlanetLab paths to iPlane destinations are one hop, we estimate that 73% of queries to* `google.com` *go directly from the client AS to Google.*

[10]Microsoft's Azure Cloud appears to block outbound traceroutes.

[11]The percentages are of total Atlas paths, not weighted.

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 do not 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 continue to shorten. 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 [9, 18]; and peering is on the rise [11, 37]. The rise of video [2] 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 as well 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 inter-domain 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. In comparison, 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 [41] using these mechanisms. However, the mechanisms generally only allow communication to neighbors: MEDs and communities usually do not propagate past one hop [36], 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,[12] but one-hop paths only consist of immediate neighbors.

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

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 [41].

Prior work proposed the joint optimization of routing between neighboring ASes [31]. Yet such protocols become more complex when they must be designed to optimize paths that traverse intermediate ASes [30], 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, alongside, or instead of 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 [6]. Second, existing approaches do not protect the AS deploying a filter, but instead prevent that AS from originating attacks on others. As a result, ASes lack strong incentives to deploy spoofing filters [6].

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 filters. 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,[13] 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,[14] 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 [28]. 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 [25], inspiring general clean-slate alternatives such as HLP [42] and simpler alternatives with restricted policies that have better convergence properties [39]. 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 [39] 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.

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

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

Avoiding outages. The Internet is susceptible to long-lasting partial outages in transit ASes [23]. The transit AS lacks visibility into end-to-end connections, so it 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 [24]. 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 [19], 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 [20], passive monitoring [4, 26], and modeling [13]. 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 [31] and CDNs and access ISPs [19], similar to approaches we imagine over direct paths. Work showing the benefits to BGP policy based only on the next hop [39] helps demonstrate the potential of such approaches. Similarly, ARROW posited that many routing problems can be lessened by tunneling to a network that offers reliable transit, and that the flattening Internet means that one tunnel is often enough [35]. Our work relies on a similar observation but an even flatter Internet.

6. CONCLUSIONS

As large content and cloud providers have been building out content distribution infrastructure and engaging in direct peering, many 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. The trend suggests that, in a departure from the current focus on general solutions, interdomain routing and traffic management techniques may benefit from optimizing for the common case of one-hop paths, a regime where simpler, deployable solutions may exist.

7. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable feedback. We also thank David Choffnes for running Sidewalk Ends AS-path conversion for our traces, Matthew Luckie for help with Scamper, Matt Calder for RIPE Atlas measurements, and Jitendra Padhye and Ratul Mahajan for helpful conversations. We thank Google for a Faculty Research Award and for a Cloud Credits Award. This work was supported in part by the National Science Foundation grants CNS-1351100 and CNS-1413978.

8. REFERENCES

[1] MaxMind GeoLite Database. `http://dev.maxmind.com/geoip/legacy/geolite/`.

[2] Sandvine global Internet phenomena report, 2014.

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

[4] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig. Web content cartography. IMC '11.

[5] 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.

[6] R. Beverly, A. Berger, Y. Hyun, and k. claffy. Understanding the efficacy of deployed Internet source address validation filtering. IMC '09.

[7] K. Butler, T. R. Farley, P. McDaniel, and J. Rexford. A survey of BGP security issues and solutions. *Proceedings of the IEEE*, January 2010.

[8] 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.

[9] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the expansion of Google's serving infrastructure. IMC '13.

[10] I. Castro, J. C. Cardona, S. Gorinsky, and P. Francois. Remote peering: More peering without Internet flattening. CoNEXT '14.

[11] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There is more to IXPs than meets the eye. *ACM SIGCOMM CCR*, October 2013.

[12] 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.

[13] A. Dhamdhere and C. Dovrolis. The Internet is flat: Modeling the transition from a transit hierarchy to a peering mesh. CoNEXT '10.

[14] B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet topology over time. *arXiv preprint*, 2012.

[15] X. Fan and J. Heidemann. Selecting representative IP addresses for Internet topology studies. IMC '10.

[16] M. T. Fangfei Chen, Ramesh K. Sitaraman. End-user mapping: Next generation request routing for content delivery. SIGCOMM '15.

[17] 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.

[18] 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.

[19] 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*, July 2013.

[20] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The flattening Internet topology: Natural evolution, unsightly barnacles or contrived collapse? PAM '08.

[21] V. Giotsas, M. Luckie, B. Huffaker, and k. claffy. Inferring complex AS relationships. IMC '14.

[22] J. P. John, E. Katz-Bassett, A. Krishnamurthy, T. Anderson, and A. Venkataramani. Consensus routing: The Internet as a distributed system. NSDI '08.

[23] 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.

[24] 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.

[25] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet routing convergence. SIGCOMM '00.

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

[27] M. Luckie. Scamper: A scalable and extensible packet prober for active measurement of the Internet. IMC '10.

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

[29] 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.

[30] R. Mahajan, D. Wetherall, and T. Anderson. Mutually controlled routing with independent ISPs. NSDI '07.

[31] R. Mahajan, D. Wetherall, and T. Anderson. Negotiation-based routing between neighboring ISPs. NSDI '05.

[32] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz. Towards an accurate AS-level Traceroute tool. SIGCOMM '03.

[33] J. Mirkovic and E. Kissel. Comparative evaluation of spoofing defenses. *IEEE Transactions on Dependable and Secure Computing*, March 2011.

[34] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (in)completeness of the observed Internet AS-level structure. *IEEE/ACM Transactions on Networking*, February 2010.

[35] S. Peter, U. Javed, Q. Zhang, D. Woos, T. Anderson, and A. Krishnamurthy. One tunnel is (often) enough. SIGCOMM '14.

[36] 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.

[37] 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.

[38] M. A. Sanchez, F. E. Bustamante, B. Krishnamurthy, W. Willinger, G. Smaragdakis, and J. Erman. Inter-domain traffic estimation for the outsider. IMC '14.

[39] M. Schapira, Y. Zhu, and J. Rexford. Putting BGP on the right path: A case for next-hop routing. HotNets '10.

[40] Y. Shavitt and U. Weinsberg. Topological trends of Internet content providers. SIMPLEX '12.

[41] N. Spring, R. Mahajan, and T. Anderson. The causes of path inflation. SIGCOMM '03.

[42] 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.

[43] 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. SIGCOMM '06.

[44] Z. Zhang, Y. Zhang, Y. C. Hu, Z. M. Mao, and R. Bush. iSPY: detecting IP prefix hijacking on my own. SIGCOMM '08.