Does Anycast Hang up on You?

Lan Wei
Information Sciences Institute
University of Southern California
Email: weilan@isi.edu

John Heidemann
Information Sciences Institute
University of Southern California
Email: johnh@isi.edu

***Abstract*—Anycast-based services today are widely used commercially, with several major providers serving thousands of important websites. However, to our knowledge, there has been only limited study of how often anycast fails because routing changes interrupt connections between users and their current anycast site. While the commercial success of anycast CDNs means anycast usually works well, do some users end up shut out of anycast? In this paper we examine data from more than 9000 geographically distributed vantage points (VPs) to 11 anycast services to evaluate this question. Our contribution is the analysis of this data to provide the first quantification of this problem, and to explore where and why it occurs. We see that about 1% of VPs are *anycast unstable*, reaching a different anycast site frequently (sometimes every query). Flips back and forth between two sites in 10 seconds are observed in selected experiments for given service and VPs. Moreover, we show that anycast instability is *persistent* for some VPs—a few VPs never see a stable connections to certain anycast services during a week or even longer. The vast majority of VPs only saw unstable routing towards one or two services instead of instability with all services, suggesting the cause of the instability lies somewhere in the path to the anycast sites. Finally, we point out that for highly-unstable VPs, their probability to hit a given site is constant, which means the flipping are happening at a fine granularity —per packet level, suggesting load balancing might be the cause to anycast routing flipping. Our findings confirm the common wisdom that anycast almost always works well, but provide evidence that a small number of locations in the Internet where specific anycast services are never stable.**

I. Introduction

A concern about anycast is that BGP routing changes can silently shift traffic from one site to another—we call this problem *potential anycast instability*. Without centralized control, such a shift will cause the connection to break. Yet this problem cannot possibly be widespread—anycast's wide use across many commercial providers suggests it works well. This observation is supported by multiple studies that have shown routing changes interrupt connections rarely [1], [10], [12]. Internet applications already must include some form of recovery from lost connections to deal with server failures and client disconnections, so anycast instability should not be a problem provided it is infrequent. Moreover, most web connections are only active for short periods of time, so the fraction of time when a route change will directly affect users is small.

In addition to BGP changes that may cause anycast instability, *load balancers* are widely used in many places in the Internet. While load balancing at the destination is usually engineered to provide stable destinations for each client, load balancing in the wide-area network is not always so careful. Prior work has observed that WAN-level load balancing can disrupt RTT estimation [14]; we believe it can also result in anycast instability. While such problems may be very rare (affecting only users that cross a specific link, and perhaps only certain traffic types), such effects in the WAN are particularly concerning because they happen *outside the control* of both the user and the service provider. It is extraordinarily difficult to detect problems that affect a tiny fraction of users, while being able to provide service to the vast majority of users. With billions of users, even a fraction of percent is a serious problem.

This paper provides the first *quantitative evaluation of the stability of anycast routing*. While very rare, we find that about 1% of combinations of vantage point and anycast service are *anycast unstable*, frequently changing routes to different sites of a service §IV-B. We call these route changes *anycast flips*, and they can disrupt anycast service by losing state shared between the VP and the server with which it was previously communicating.

This result follows from the study of 11 different anycast deployments, each a global *Root DNS Letter* with an independent architecture, with sizes varying from 5 to about 150 anycast sites (each a location with its own anycast catchment and one or more servers).

Our second contribution is to demonstrate the *severity and potential causes of* of route flips through a number of measurement studies. This study provides a *broad* view by examining all combinations of about 9000 VPs and 11 anycast services. We use several measurement methods to examine how frequent flips are, proving they often flip between anycast sites in tens of seconds §IV-E, and strongly suggesting they may flip more frequently, perhaps every packet §IV-F. We also find that anycast instability is often continuous and persistent: 80% of unstable pairs of VP and anycast services are unstable for more than a week. For a few, these problems are very long lasting: 15% VPs are still unstable with some service even 8 months later §IV-C. We show that *anycast instability is specific to paths*: with almost all VPs with instability seeing it in only a few services (one to three), not all 11 §IV-D. Although we cannot definitively know the root causes of anycast instability, we do show from our measurement that certain VP/service pairs keeps flipping very frequently, likely every packet §IV-F. A possible explanation is load balancers on WAN links.

Our results have two important implications. First, *anycast*

almost always works without routing problems: for 99% of combinations of VP and anycast service, routes are stable for hours, days, or longer. With multiple successful commercial CDNs, this result is not surprising, but it is still important to quantify it with a clear, public experimental. Second, we show that *anycast does not work for **all** locations, and a few VP/anycast combinations (about 1%) see persistently route instabilities* with paths flipping frequently and perhaps every packet. This result suggests that commercial anycast CDNs and those providers who want to provide service to *all* users may wish to study locations that have anycast unstable routes and investigate ways to reduce this instability.

II. Anycast Routing Instability

In IP anycast, an anycast service uses a single IP address, and a user's traffic is directed to a "nearby" site selected by BGP routing. Typically, "nearby" is defined by the length of the path in AS hops, but BGP supports multiple mechanisms that allow service operators and ISPs to impose policy decisions on routing (for details, see an overview [3]). Policies can be political (this path is only for academic traffic), commercial (send more traffic to the less expensive peer), or technical (load balance across these links).

Anycast flips can be a problem because changes in routing shift a client to a new server without any notification to either. If the client and server have some shared states, such as active TCP connections, these will break because of a TCP reset and need to be restarted.

CDNs often keep persistent TCP connections open to clients when sending streaming media such as video. While applications need to be prepared for unexpected termination of TCP connections, a route flip will greatly increase latency as the problem is discovered and a new connection is built.

Most DNS today is sent over UDP, but zone transfers use TCP, and recent work has suggested widespread use of TCP and TLS for DNS privacy [22], [11]. For DNS, a route flip results in a much larger response time. For a CDN or video streaming, it might result in playback stalls and "buffering" messages.

A key factor affecting the degree of impact that an anycast flip has on the client is how long its TCP connections are open, plus how long they are active. For video, connections may be open for many tens of minutes, during which they may be active around 10% of the time. For DNS, connections may be open for tens of seconds and active briefly, but multiple times.

III. Methodology

This section explains the essential features of the datasets and the methodology we use to analyze the dataset.

III-A Sources and Targets

Our paper uses three CHAOS query datasets listed in Table I. All use the RIPE Atlas infrastructure [15]. Two are existing public datasets they collect [16], the third is an additional publicly-available dataset we collect to improve time precision [17]. Additionally, we use RouteViews dataset [19] to check the updates of BGP routing table.

start	**duration**	**VPs number**	**probing interval**	**targets** (root letters)
2015-12-05 00:00	7 days	9184	240s	`--CDEFG-IJKLM`
2016-08-01 00:00	7 days	9254	240s	`A-CDEFG-IJKLM`
2017-01-29 21:00	30 minutes	100	20s	`---D---------`
2016-08-01 00:00	7 days	192 peers	cont.	`A-CDEFG-IJKLM`

TABLE I: Datasets observing catchments from UDP-based CHAOS queries, and BGP routing updates.

The target of RIPE data collection is all 13 Root DNS Name Servers (or *Root Letters*), shown in Table II. Of these services, our study considers all Root Letters that use anycast at the time of measurement. We omit A-Root from the 2015 dataset, because at that time it was only probed every 30 minutes. We omit B- and H-Root from both datasets because, at these times, B is unicast and H uses primary/secondary routing. Root Letters are operated by 12 organizations and use 13 different deployment architectures, and a wide range of sites (5 to 144), providing a diverse set of targets.

We actively probe each anycast letter, sending queries from more than 9000 RIPE Atlas probes, embedded computers we call *Vantage Points* (VPs). VPs are geographically distributed around the world, although North Africa and China are only sparsely instrumented. Our results may underrepresent anycast problems in these two areas.

We use active queries rather than passive analysis of BGP because we are most concerned about frequent flipping (§IV-E) and prior work has shown that BGP changes are relatively infrequent, often hours or more apart [13]. We also use BGP data from RouteViews to confirm the BGP stability (§IV-F).

The targets of our queries are 11 root-letters which are operated by 10 organizations, listed in Table II. Different letters have different deployment architectures, and they have from 5 to 144 anycast sites (a wide range). Although we study most letters, we do not see all anycast sites of each letter. We sometimes miss sites because RIPE Atlas VPs are sparse in some parts the world (particularly Africa), and because some anycast sites are local-only and so will be seen only by a VP in the same AS. Fortunately, answers to our research questions do not require complete coverage.

We do not directly study anycast-based CDNs. Like root letters [18], CDNs vary widely in size, from ten or tens of sites [2], even approaching 1000 sites [4], although hybrid architectures may use a subset of all sites [8]. In §IV-D we show that instability often results in the network, so our results likely apply to CDNs.

III-B Queries from RIPE Atlas

Each VP queries each Root letter every 4 minutes (except for A root in the 2015 dataset). The query is a DNS CHAOS class, for a TXT record with name `hostname.bind`; this query is standardized to report a string determined by the server administrator that identifies server and site [21]. Queries are directed at anycast IP addresses that are served by a

letter	operator	sites reported	observed 2015	observed 2016
A	Verisign	5	—	5
C	Cogent	8	8	8
D	U. Maryland	87	63	71
E	NASA	71	74	66
F	ISC	59	51	48
G	U.S. DoD	6	6	5
I	Netnod	49	51	56
J	Verisign	98	65	89
K	RIPE	33	32	40
L	ICANN	144	110	118
M	WIDE	7	6	6

TABLE II: Targets of our study are most of 13 Root Letters, with their reported number of sites [18], and how many sites we observe in each datasets.

specific Root Letter. They do not follow usual DNS selection rules, allowing us to study each letter as an independent service.

The above query results in a record listing the time, the VP's identity, and the response to the CHAOS query (or an error code if there is no valid response). The responses are unique to each server in use. There is nothing to prevent third parties from responding on an anycast service address, and we see evidence of that in our data. We call responses by third parties other than the operator *spoofed*.

We map the CHAOS responses we see to the list of sites each self-reports [18], following practices in prior studies [7]. While CHAOS responses are not standardized, most letters follow regular patterns, and the same pattern from many different VPs gives us some confidence that it is valid. For example, if lax1a.c.root-servers.org and lax1b.c.root-servers.org regularly appear, we assume city.c.root-servers.org is C-Root's pattern.

CHAOS responses usually identify specific servers, not sites. Some letters have multiple servers at a given site. Continuing the above example, lax1a.c.root-servers.org and lax1b.c.root-servers.org suggest C-root has two servers 1a 1b at the lax site. Not all letters identify servers inside large sites, but all provide unique per-site responses.

We study flipping between *sites* and ignore changes between *servers* in each site, since operators can control server selection if they desire (perhaps with a stateful or consistent load balancer), but not changes between sites.

We detect spoofed strings as those that do not follow the pattern shown by that letter, those seen only from a few VPs in specific networks, and because spoofed replies typically reply with very low latency (a few ms instead of tens of ms). Typically, about 0.7% of VPs see spoofed replies, and those VPs *always* see the same replies, suggesting their ISPs intercept DNS. While we work to remove spoofed addresses from our data, our methods do not prevent a malicious party from generating correct-looking replies.

III-C Other Sources: High Precision Queries and BGP

In addition to the standard RIPE Atlas probes of Root letters, we also request our own measurements at a more frequent time interval and gathered BGP information to understand routing.

For high precision queries we use the RIPE Atlas infrastructure, but select 100 VPs of interest, based on those that see anycast instability. For these VPs, we request that they query D-Root every 60, 70, 80, and 90 s for 30 minutes. Although RIPE Atlas limits queries to once per minute, by scheduling concurrent measurement tasks on the same VPs we can get results that provide precision approaching 20 s or even less from the unevenly-distributed queries.

To rule out BGP as the cause of flipping we use data from RouteViews [19] from 2016-08-01 to 2016-08-07. Although the peers that provide routing data are in different locations than our RIPE VPs, the multiple RouteViews peers provide a guiding picture of Internet routing for BGP.

III-D Detecting Routing Flips

We define a *routing flip* as when a prior response for a VP's query indicates one site, and the next response indicates a different site. For missing replies, we assume that the VP is still associated with the same site as in the prior successful reply.

Most VPs miss ten or fewer replies per day and so loss does not change our results. About 200 VPs (around 2%) miss all or nearly all replies; we exclude these from our datasets.

IV. Evaluation

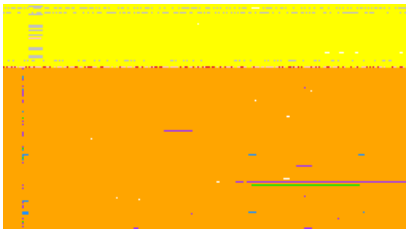
We next apply our analysis to evaluate anycast stability. We first identify examples of routing stability, then quantify how often it happens, how long it persists, and then discuss possible causes for the instability.

IV-A What Does Anycast Instability Look Like?

We first look to see if there is *any* anycast instability. While successful anycast-based CDNs suggest that most users will be stable, but perhaps a few are less fortunate. We look at the data from RIPE Atlas to the Roots Letters as described in §III, looking for VPs that change sites between consecutive queries.

Before looking at stability statistics, we first show a sample of direct observations to characterize typical anycast stability. We selected 140 VPs from the dataset for C-Root in 2015 dataset and plotted which sites they access for each 40-minute period of the week-long dataset. To better present the data, we select C-root because we can assign each of its 6 sites a unique color (or shade of gray). We choose 140 VPs that mainly associate with the MAD and ORD sites as representative of all sites for C. To show our full week of data on the page, we report only the last site selected by each VP in each 40 minute period. (This summarization actually reduces the apparent amount of changes.)

Figure 1 is a timeseries showing which sites 140 VPs reach over 1 week in the 2015 dataset. Each row is a VP, and each column is a 40-minute period, and color indicates the currently active catchment (or white if no reply). Figure 1b zooms in



(a) 140 VPs for 1 week

(b) The top left: 50 VPs for 20 hours

Fig. 1: Sites accessed by 140 VPs: each row represents a VP, and each column represents a 40-minute period, and the colors show what site that VP reaches for C-Root (yellow: MAD, orange: ORD, gray: CDG, red: BTS, pink: FRA, purple: IAD, blue: JFK, green: LAX, white: no response). Dataset: 2015.

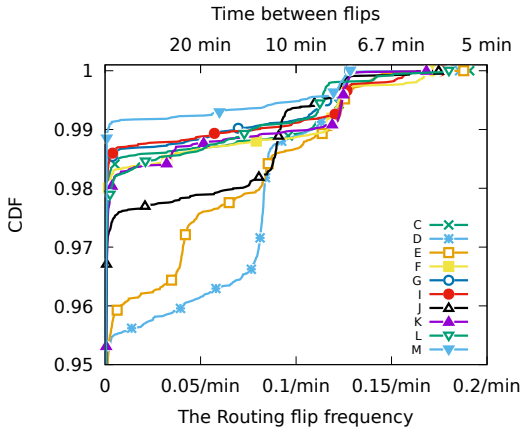
on the top left 50 VPs for 20 hours. We see similar results for other letters, and for other datasets.

Overall stability: Figure 1 shows very strongly that anycast usually works well—*most VPs are very stable*. Many of these VPs access one site, with most of top 39 VPs for MAD (the yellow band), while the most of the bottom 101 VPs for ORD (orange). We expect general stability, consistency with wide, successful use of anycast.

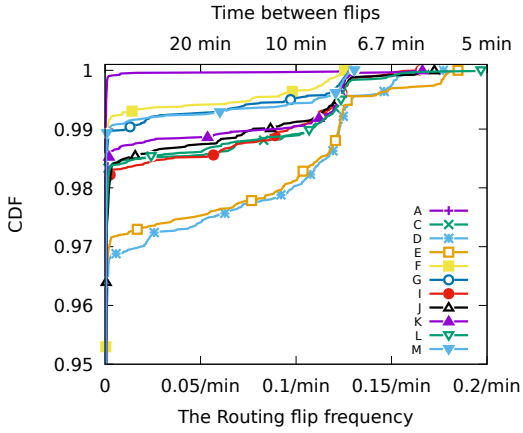
While most VPs are stable, we next look at three groups of routing flips as shown by color changes in the figure.

Groups of Routing Flips: These routing changes happen and affect many VPs at the same time; if these are occasional they are benign. On the left of Figure 1a, there is a tall vertical "stripe" affecting many VPs for ORD (orange), and another wider strip affecting many of the VPs for MAD (yellow). In each of these cases we believe there was a change in routing in the middle of the network that affected many (but not all) users, changing them from ORD or MAD to blue JFK. In both cases, the routes changed back fairly quickly (after 36 minutes for MAD-CDG-MAD, and 6 hours for ORD-LAX/IAD/JFK-ORD). Group flips that happen occasionally will require TCP connection restarts, but two events in two weeks will have minimal impact on users. These kind of normal routing changes reflect anycast automatically re-routing as ISPs reconfigure due to traffic shifts or link maintenance.

Individual, Long-term Changes: Other times we see individual VPs change their active site, perhaps reducing latency. For example, the bottom-right of Figure 1a, about 10 VPs change from ORD to IAD or LAX (orange to purple or green), and stay at that site for the remainder of the period, about three days. Again, we believe these changes in routing represent long-term shifts in the network, studied elsewhere [20]. Because these changes are infrequent, long-term shifts cause minimal harm to users, and sometimes they may help if they result in a lower latency path. They may also represent routing changes by operators to re-balance load on sites.



(a) Week of 2015-12-05.



(b) Week of 2016-08-01.

Fig. 2: Cumulative distribution of mean flip time for each VP, broken down by anycast service. (Note the y-axis does not start at zero.)

Frequent Routing Flips Finally, we see a few cases where *VPs see persistent routing flips*, suggesting that, *for them, anycast will not work well*. In Figure 1a we see four cases: three VPs flip between MAD and CDG and back (gray and yellow, all in the yellow band), and one VP alternates between FRA-BTS-MAD (pink, red and yellow, shown at the boundary of the yellow and orange bands). This behavior continues throughout the week. While the VPs sometimes reach the same site in consecutive measurements, this kind of frequent flipping greatly increases the chances of breaking TCP connections.

IV-B Is Anycast Instability Long Lasting, and for How Many?

We have seen some unstable users (Figure 1a), but *how many* are unstable? To answer that question, we must first consider how long instability lasts.

To evaluate the stability of each VP, we compute the mean duration that VP is at each site, then report the cumulative distribution for each root letter for 2015 and 2016 dataset in Figure 2.

Root Letter	mean	(sd)	flips (% VPs) =0	≤ 1	≤ 2	≤ 3
A	2.0	(21.2)	23%	25%	98%	98%
C	16.7	(133.2)	80%	80%	90%	91%
D	32.4	(188.5)	50%	52%	89%	91%
E	30.9	(190.0)	66%	69%	90%	90%
F	7.1	(81.8)	81%	82%	91%	92%
G	11.3	(93.5)	12%	12%	51%	52%
I	17.2	(134.3)	72%	76%	89%	90%
J	15.6	(128.3)	69%	72%	90%	92%
K	14.5	(124.8)	76%	78%	86%	86%
L	17.1	(137.8)	71%	75%	90%	92%
M	8.9	(98.1)	90%	91%	95%	95%

TABLE III: Number of flips per VP, for each Root Letter, for the week of 2016-08-01.

The result confirms the prior observation that *overall, anycast is very stable for most VPs*. The y-axis of the CDFs (Figure 2) does not start at zero, and we see that 90% of VPs see two for fewer changes for all Root Letters we study but one (Table III). In fact, A-root **barely** saw any route changes for any VPs in the week starting from 2016-08-01. (Figure 2b).

Stability means *most VPs are in one catchment for a long time*. Table III shows overall statistics per letter, for each dataset. Most VPs are very stable.

However, it also confirms *a few VPs experience frequent routing flips*. We define a VP as *anycast unstable* when the mean time between flips is 10 minutes or less. We select this threshold because it is slightly longer than two measurement intervals (each 4 minutes), tolerating some measuring jitter. Based on the threshold of 10 minutes, we see that about 1% of VPs are anycast unstable for almost all Root Letters for both 2015 and 2016 datasets. One exception is A-root, who shows high stability in the 2016 dataset. This analysis suggests that, at least in these datasets, some VPs will have a difficult time using anycast and may experience TCP connection breaks.

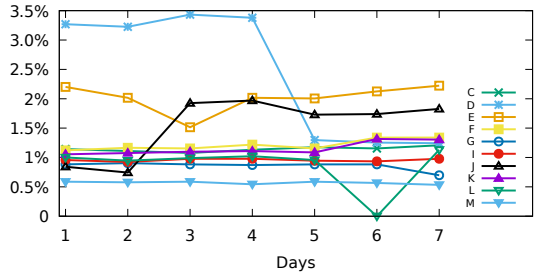
To confirm these results are typical, Figure 3 examines the fraction of anycast unstable VPs each day. Most VPs consistently have about 1% of VPs as unstable, although there is some variation in a few letters (for example, D has 3.2% in part of Figure 3a).

The precision of results in Figure 2 is limited by the 4 minute frequency of basic RIPE observations. We later return to this question with more frequent request rate and analysis to suggest that actually flipping rates are much higher than every 4 minutes (§IV-E), and likely every packet (§IV-F).

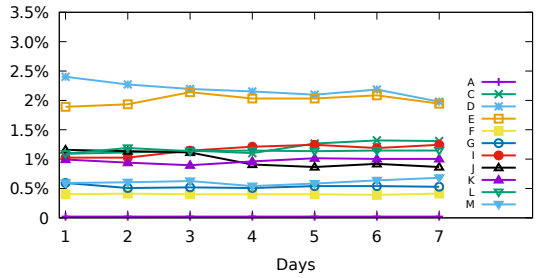
IV-C Is Anycast Instability Persistent for a User?

We have shown that about 1% VPs are anycast unstable, and that this count is relatively consistent over time (Figure 3). But does instability haunt specific users, or does it shift from user to user over time? That is: is the *set of unstable users* itself stable or changing?

To evaluate if anycast instability is persistent, we split each week into its first half and second half. We identify anycast unstable VPs in each half using our 10 minute threshold, then we compare the two sets to see how much overlap they have.



(a) Week of 2015-12-05.



(b) Week of 2016-08-01.

Fig. 3: The percentage of anycast unstable VPs for each day in a week.

Table IV shows the number of unstable VPs in each half of the week, for both datasets. While the absolute number of unstable VPs varies by letter, the *most VPs that are unstable keep being unstable over the whole week*—the percent that overlap in the two halves of the week is at least 63% and typically around 90%. Anycast instability is a stable property between a VP and its anycast service. Although the two weeks we checked are more than half a year apart, we check the overlap over two different weeks and found there are still around 13% overlap.

It is also possible we see large amounts of overlap because many VPs are on the same networks—we rule this case out with additional validation. To check for bias from clustered VPs, we manually examined unstable VPs and their ISPs. We found they were dispersed across many different networks and had no similar /24 prefixes.

This analysis shows unlucky VPs (those that are anycast unstable) are likely to continue to be unlucky. This result suggests that we must take care in interpreting the commercial success of anycast CDNs. Although they work well for most users, and analysis of their own data shows few broken TCP connections, it may be that their sample is not abundant enough, especially the VPs' coverage, because we just showed the instability is sticky with specific VPs over time. People will use anycast CDNs that work, but unlucky people that are anycast unstable for a particular CDN may simply turn away from that CDN (or its clients) because it doesn't "work" for them.

Root Letter	week of 2015-12-05 unstable VPs 1st	2nd	both	Overlap (percent)	week of 2016-08-01 unstable VPs 1st	2nd	both	Overlap (percent)	both weeks Overlap (percent)
A	—	—	—	—	2	2	2	100%	—
C	102	108	98	93%	97	113	67	64%	10%
D	301	106	99	63%	190	186	142	75%	14%
E	119	180	110	76%	173	183	129	72%	13%
F	107	111	107	98%	34	35	26	75%	7%
G	82	82	76	92%	44	49	30	64%	16%
I	84	85	76	89%	84	107	68	72%	9%
J	68	157	48	50%	94	74	64	77%	12%
K	99	100	94	94%	86	93	75	89%	18%
L	87	67	62	81%	93	102	80	82%	20%
M	53	52	46	87%	55	57	32	57%	24%

TABLE IV: Overlap of anycast instability for specific VPs in half-weeks.

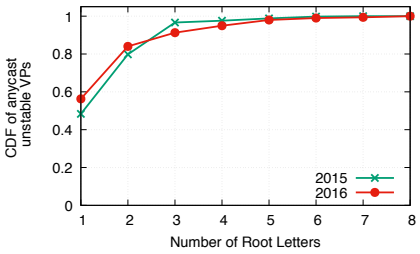


Fig. 4: The CDF of unstable VPs for how many root DNS services. The vast majority of VPs only experience instability towards one to three services.

IV-D Is Anycast Instability Near the Client?

We next look at *where in the network* anycast instability appears to originate. Is it near the VP (the client), or near the anycast service's sites (the servers)? This question is of critical importance, because we have shown that some VPs are consistently anycast unstable. If the problem is located near the VP, it is likely that they will be unstable with *many* anycast services, and if an important service (like a CDN or Root DNS services) is provided *only* by anycast, then it might impossible for that VP to get service.

To explore this question, we use the same approach we used to study the persistence of anycast instability (§IV-C), but rather than comparing two halves of the same week, we compare different anycast services (different Root Letters). We consider three cases: (1) If instability was near a specific anycast site, then *many* VPs reaching that site should see instability. (2) If anycast instability is near a specific VP, we expect that that VP will be unstable with *many* anycast services. (3) On the other hand, if a VP is unstable with only one service, then the problem is likely in the middle of the network on the particular path from that VP to its current site.

We rule out case (1), since the service operator would notice and correct a site-specific problem. In addition, our study of number of unstable VPs per service that showed that there are at most a few anycast unstable VPs for each service (Figure 2).

For the 2015 dataset we identify 416 VPs that are anycast unstable for some root letter. Anycast instability is a property between the VP and a specific service, and Figure 4 shows for how many anycast services each of these VPs find to be unstable.

Our first observation is that *almost half of VPs are only unstable with one service*. Of the 416 VPs, 200 (48%) are unstable with only one of the 11 IP anycast services we study. We conclude that *the most common location of anycast instability is the middle of the network*, somewhere on a unique network path, not near the VP or an anycast site.

About the same number are anycast unstable with two or three services—202 of the 416 VPs, again about 48%. We conjecture that in these cases the problem is closer to the VP. Fortunately, it does not affect all services.

Only 2% of VPs are anycast unstable with more than 3 services, and none are unstable with more than 7. Since very few VPs have problems with all anycast services, we rule out case (2), that there are problems that are not very near the VPs.

The distribution of 2016 dataset is similar to 2015. The new weeks saw 494 unstable VPs, more than the 2015 datasets. The fact that highest number of letters the VP experiencing instability simultaneously goes from 8 to 7 is normal considering we add another letter in our dataset.

One source of instability are paths that are load balanced over multiple links, where link selection is a function of information in packets that change in each packet. For example, the UDP source port is randomized in each of our queries; if it is included in a hash function for load balancing, packets could take different links. This problem has previously been observed in ping-based latency measurements [14]. Additional work is provided in following sections §IV-F trying to understand these root causes.

We do not consider correlations between number of sites and degree of flipping. One might look at Figure 2 for correlations, but with only 11 architectures, each unique, it seems difficult to make statistically strong comparisons.

We conclude that anycast instability is not near the VPs, nor the anycast sites, but a factor of the path between them, depending on their relative locations. The good news of this conclusion is that it means clients that see problems with one anycast service can simply try a different one. That implies anycast instability is unlikely to be a problem for Root DNS service (with a dozen independently architected

anycast services), and it may be a concern for CDNs that operate anycast services [8].

IV-E Higher Precision Probing Shows More Frequent Flipping

The long-term RIPE Atlas datasets (examined in §IV-A) provide broad coverage for years, but each VP observes its catchment every 4 minutes, and we would like greater precision on flip frequency. We use this data to identify anycast unstable VP/service pairs, but these measurements are hugely *undersampled*—we expect some sites are flipping every packet, but 4 minute measurements of a VP flipping between two sites every packet will see a median flip time of 8 minutes. Improving the precision of this estimation is important because TCP connections are active for short times, often few tens of seconds, so proof of 4 minute flipping does not demonstrate TCP problems. In this section we take additional, direct measurements from RIPE Atlas to evaluate if these pairs are actually flipping more frequently than standard RIPE Atlas measurements are able to observe. (We cannot run custom measurement code on the VPs because RIPE does not support that, we have no way of contacting VP owners, and we require data from the few, specific VPs that show frequent flipping.)

To test this question we select 100 VPs to probe towards D-root in 30 minutes with unevenly distributed 95 probes, roughly one query per 20 seconds.

Figure 5 compares how many flips we see when the same VPs probe at 4 minute intervals (green filled dots on the bottom) compared to probes sent with about 20 s intervals (top open squares), for these 100 VPs with frequent flips. (We report counts of flips rather than mean flip duration because it is difficult to assess mean duration with this hour-long measurement.)

This data shows that more observations result in more flips—the open squares are always above filled dots. Of course more observations make more flips possible, but this data shows that 4 minutes measurements are *undersampled* and the path is flipping much more often. In fact, two VPs marked with asterisks show no flips during 4 minute observations, even though they flip frequently about at least every 30 seconds. If we assume every packet flips, then with fewer samples, these VPs just get "lucky" and appear stable with undersampling.

Since our measurements are not synchronized, sometimes we take measurements very close in time. As two specific examples, we saw one VP (84.246.12.69) flip from London to Frankfurt and back with three measurements in 7 s, and another (201.217.128.115) flip from Miami to Virginia and back in 10 s. In the next section we provide statistic evidences that suggest per-packet flipping are happening for specific VPs.

IV-F Does Per-Packet Flipping Occur?

We have shown that some VPs see very frequent flipping to some anycast services—as short as tens of seconds (§IV-E). It seems unlikely that BGP is changing so frequently, since route flap damping is usually configured to suppress multiple changes within a few minutes.

To rule out BGP as the source of instability we analyse the dataset of RouteViews [19] from 2016-08-01 to 2016-08-07, finding that BGP is quite stable. Of the total 192 BGP RouteViews peers we studied, 0% to 23% RouteViews peers will ever see a BGP change, for each root letter, on each day. The mean time of BGP change seen by those RouteViews peers is always fewer than twice a day, so such infrequent routing changes cannot explain anycast flips that occur multiple times per minute. This new analysis supports previous studies that show BGP changes are relatively infrequent for Root DNS [13]. Instead, we suggest that these very frequent flips result from per-packet decisions made by load balancers in the path.

We cannot *directly* evaluate very frequent flips, because they occur only from specific VPs to certain anycast services. While we find them with RIPE Atlas, it limits probing intervals to 60 s, and even with multiple concurrent experiments, sub-second probing is impossible on RIPE. Neither can we reproduce these flips from another site, since they are specific to the path from that VP to the service.

However, we can *indirectly* show it is likely that these flips are per-packet by looking at how they respond in sliding time window over time. If the path is flipping every packet, then the probability of reaching a specific site should be *almost consistent over time*. We measure consistency by sliding a window over all observations and looking at the fraction of queries that go to different anycast sites. If flipping is per-packet, the fraction should be similar for any window duration and time period.

To evaluate this hypothesis, we return to the 2016 dataset, and we focus on the 100 VPs that have frequent site-slips towards C-Root (measured as those with time-to-flip around 10 minutes, roughly twice the measurement frequency). For each VP, we compute how many times their requests reached a specific anycast site in a window of 20 observations. We slide the window forward with one observation at a time, so windows overlap.

Figure 6 shows a representative example for one VP (146.186.115.74), which flips between the JFK and ORD sites of C-Root. We report the fraction of time the VP is at JFK, measured with a 20-observation moving window. We compute this moving window at three timescales, first using all the data (4 minute samples, the wide line), and also downsampled two times (8 and 16 minutes). First, we see that the long-term average is around 0.5, consistent with each packet going one way or the other. There are peaks and valleys, as we expect with any long-term average, sometimes we get a run of one site or the other, but standard deviations is ± 0.1184 (shown as the dashed lines), and most of the time the average is within this range. However, lack of any repeating pattern suggests that there are not long-term flips, but per-packet.

In addition, when we compare the three timescales, all show similar properties. This result is consistent with all being drawn from random samples of per-packet flipping. These

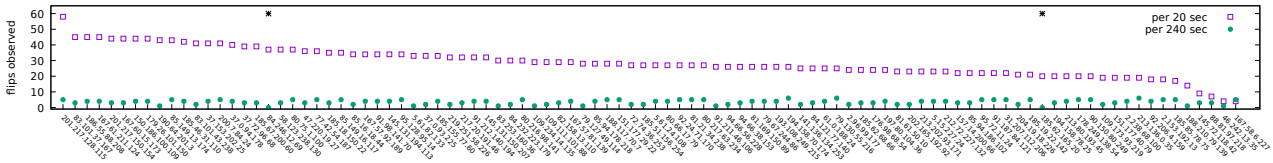


Fig. 5: Counting site flips from 100 VPs to D-Root. Measurements with about 20 s intervals (blue open squares on top) are compared to every 4 minutes (green filled dots on bottom). Two VPs with no flips in 4 minute data are marked with an asterisk (*).

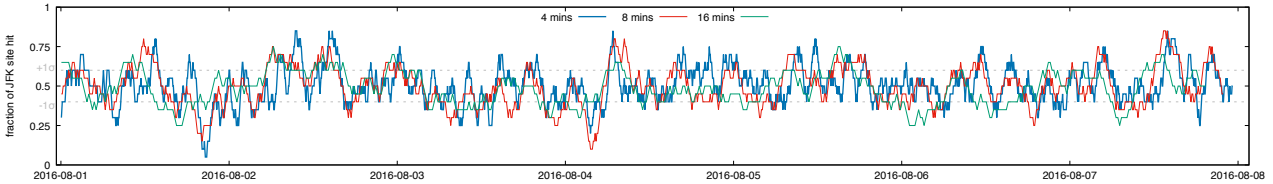


Fig. 6: Fraction of time one VP (146.186.115.74) spends at the JFK site of C-Root. Each point is the mean of a 20-observation sliding window, done at four timescales, 4 minutes (wide blue), 8 minutes (red), and 16 minutes (blue).

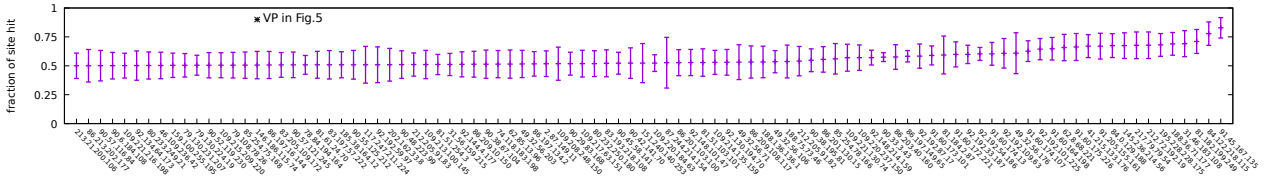


Fig. 7: Mean and standard deviation of site hit ratio across all sliding time window in a week.

trends supports the suggestion that we would see similar results if we increase sampling frequency, as we showed experimentally to 20 s in §IV-E.

We see the same behavior for this example VP in most of the other 100 VPs we observed. Figure 7 shows the mean and standard deviation of all selected VPs, sorted by mean. Most of these VPs show a ratio around 0.5 and a standard deviation around 0.1, consistent with our example, and consistent with random selection per-packet. Some sites on the right of the graph show an uneven split; we expect these are due to uneven load balancing, or multiple load-balanced paths.

Taken together, our experiments and this analysis present a strong case for per-packet flipping. Experiments at 4 minutes and 20 s (§IV-E) directly support this claim, and our analysis at multiple timescales and across many VPs indirectly supports it.

IV-G Open Question: Per-Packet Or Per-Flow for TCP?

Load balancers typically operate on flows, not packets, so it is possible the packet-flipping we observe in UDP will not affect TCP connections. RIPE Atlas partially supports TCP-based DNS queries. As of April 2017, these TCP queries do not work on most RIPE VPs (due to a limit in firmware version 4760), but it works on the 250 RIPE Anchors that run firmware 4770. We tested TCP-based DNS queries on these 250 anchors, but frequent flipping is already quite rare, and anchors are intentionally in "more stable" locations (according RIPE). We found only 37 (Anchor, letter) combinations that showed frequent UDP flipping, and 11 combinations where TCP consistently failed, but there was no overlap. We plan to evaluate TCP queries on all 9000 VPs after their firmware has been upgraded.

V. Related Work

Prior studies have considered many aspects of anycast: latency [5], [20], geography [1], usage and traffic characteristics [10], [7], [6], [9], CDN load balancing [8], and performance under DDoS attack[13]. However, only a few studies have considered the stability of anycast [12], [1], and their conclusions are largely qualitative. Unlike this prior work, our goal is to *quantify the stability* of anycast.

Direct measurements: Prior stability studies either directly or indirectly measured catchments. Direct measurement studies of anycast stability use data from end-users or monitors that contact the anycast site. Microsoft has used Bing clients to study anycast and evaluate latency and load balancing. They observed 21% of end-users change sites at least once per a week [5]. However, the FastRoute system is concerned about small file downloads in their availability studies [8]. They also showed that anycast availability dipped from 99.9% to 99.6% once during their week-long observation, but do not discuss why.

LinkedIn[1] evaluated anycast with a synthetic monitoring service to evaluate latency and instability, and did not find "substantial instability problems". Our results suggest that most VPs are stable, so long-duration observation is unlikely to see new results, unless one studies from more vantage points located at other different places in the Internet.

Finally, recent studies of DNS Root anycast showed frequent routing flips during DDOS [13], but that paper did not study stability during normal periods.

Our work is also direct measurement like these prior studies, but unlike prior work we use many geographically dispersed VPs (more than 9000 from RIPE Atlas) multiple services (the 11 anycast Root DNS services, some with 100 sites), under normal behavior.

Indirect evaluation: Inference can estimate changes in anycast catchments by looking for changes in latency or hop counts (IP time-to-live). Cicalese and Giordano examined anycast CDN traffic by actively sending queries to each prefixes announced by 8 CDN providers [6]. The found anycast stable, with nearly constant RTT and time-to-first-byte, and consistent TTLs over a month. They later studied the duration of TCP connections for DNS and show that most last tens of seconds, suggesting that DNS will not be affected by infrequent anycast catchment changes [9]. Unlike their work, we directly observe site flips with CHAOS queries, rather than infer it. More important, we use 9000 VPs geographically dispersed across the world, while their study is based on VPs only in Europe.

VI. Conclusion

In this paper we used data from more than 9000 vantage points (VPs) to study 11 anycast services to examine the stability of site selection. Consistent with wide use of anycast in CDNs, we found that anycast almost always works—98% of VPs see few or no changes. However, we found a few VPs—about 1%—that see frequent route changes and so are *anycast unstable*. We showed that anycast instability in these VPs is usually "sticky", persisting over a week of study. The fortunate fact, that most unstable VPs are only affected by one or two services, shows instability causes may lie somewhere in the middle of the routing path. By launching more frequent requests, we captured very frequent (back and forth within 10s) routing change in our experiments using the unstable VPs we discovered from previous analysis, the statistical analysis shows they are affected by per-packet flipping, which is potentially caused by load balancer in the path. Our results confirm that anycast generally works well, but when it comes to a specific service, there might be a few users experiencing routing that are never stable.

Acknowledgments: This research has been partially supported by measurements obtained from RIPE Atlas, an open measurements platform operated by RIPE NCC. We thank them for sharing their data.

Lan Wei and John Heidemann's work is partially sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, HSARPA, Cyber Security Division, BAA 11-01-RIKA and Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0344 and via contract number HHSP233201600010C. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views contained herein are those of the authors and do not necessarily represent those of DHS or the U.S. Government.

References

[1] P. Bret, K. Prashanth, J. Samir, and A. K. Zaid. TCP over IP anycast—pipe dream or reality? https://engineering.linkedin.com/network-performance/tcp-over-ip-anycast-pipe-dream-or-reality, Sept. 2010.

[2] CacheFly Network Map. *https://web1.cachefly.net/assets/network-map.html*, Apr. 2017.

[3] M. Caesar and J. Rexford. BGP routing policies in ISP networks. *IEEE Network Magazine*, 19(6):5–11, Nov. 2005.

[4] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the expansion of Google's serving infrastructure. In *Proceedings of the ACM Internet Measurement Conference*, pages 313–326, Barcelona, Spain, Oct. 2013. ACM.

[5] M. Calder, A. Flavel, E. Katz-Bassett, R. Mahajan, and J. Padhye. Analyzing the performance of an anycast cdn. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 531–537. ACM, 2015.

[6] D. Cicalese, D. Giordano, A. Finamore, M. Mellia, M. Munafò, D. Rossi, and D. Joumblatt. A first look at anycast cdn traffic. *arXiv preprint arXiv:1505.00946*, 2015.

[7] X. Fan, J. Heidemann, and R. Govindan. Evaluating anycast in the domain name system. In *INFOCOM, 2013 Proceedings IEEE*, pages 1681–1689. IEEE, 2013.

[8] A. Flavel, P. Mani, D. A. Maltz, N. Holt, J. Liu, Y. Chen, and O. Surmachev. FastRoute: A scalable load-aware anycast routing architecture for modern CDNs. In *Proceedings of the USENIX Symposium on Network Systems Design and Implementation*, Oakland, CA, USA, May 2015. USENIX.

[9] D. Giordano, D. Cicalese, A. Finamore, M. Mellia, M. Munafò, D. Z. Joumblatt, and D. Rossi. A first characterization of anycast traffic from passive traces. In *Proceedings of the IFIP Traffic Monitoring and Analysis Workshop (TMA)*, 2016.

[10] J. Hiebert, P. Boothe, R. Bush, and L. Lynch. Determining the cause and frequency of routing instability with anycast. In *Proceedings of the Asian Internet Engineering Conference (AINTEC)*, pages 172–185, Pathumthani, Thailand, Nov. 2006. Springer-Verlag.

[11] Z. Hu, L. Zhu, J. Heidemann, A. Mankin, D. Wessels, and P. Hoffman. Specification for DNS over Transport Layer Security (TLS). RFC 7858, Internet Request For Comments, May 2016.

[12] M. Levine, B. Lyon, and T. Underwood. TCP anycast—don't believe the FUD. Presentation at NANOG 37, June 2006.

[13] G. C. M. Moura, R. de O. Schmidt, J. Heidemann, W. B. de Vries, M. Müller, L. Wei, and C. Hesselman. Anycast vs. DDoS: Evaluating the November 2015 root DNS event. In *Proceedings of the ACM Internet Measurement Conference*, Nov. 2016.

[14] C. Pelsser, L. Cittadini, S. Vissicchio, and R. Bush. From Paris to Tokyo: On the suitability of ping to measure latency. In *Proceedings of the ACM Internet Measurement Conference*, Barcelona, Spain, Oct. 2013. ACM.

[15] RIPE NCC. DNSMON. https://atlas.ripe.net/dnsmon/, 2015.

[16] RIPE NCC. RIPE Atlas root server data. https://atlas.ripe.net/measurements/ID, 2015. ID is the per-root-letter experiment ID: A: 10309, B: 10310, C: 10311, D: 10312, E: 10313, F:10304, G: 10314, H: 10315, I: 10305, J: 10316, K: 10301, L: 10308, M: 10306.

[17] RIPE NCC. RIPE Atlas self measurement. https://atlas.ripe.net/measurements/ID, 2017. ID is the per-frequency experiment ID: 60s : 7788714, 70s : 7788717, 80s: 7788718, 90s:7788729.

[18] Root Operators. *http://www.root-servers.org*, Apr. 2016.

[19] RouteViews. routeviews2, routeviews3, routeviews4. http://bgpmon.io/archive/help, Aug. 2016.

[20] R. d. O. Schmidt, J. Heidemann, and J. H. Kuipers. Anycast latency: How many sites are enough? In *Proceedings of the Passive and Active Measurement Workshop*, page to appear, Sydney, Australia, May 2017. Springer.

[21] S. Woolf and D. Conrad. Requirements for a mechanism identifying a name server instance. RFC 4892, Internet Request For Comments, June 2007.

[22] L. Zhu, Z. Hu, J. Heidemann, D. Wessels, A. Mankin, and N. Somaiya. Connection-oriented DNS to improve privacy and security. In *Proceedings of the 36th IEEE Symposium on Security and Privacy*, pages 171–186, San Jose, Californa, USA, May 2015. IEEE.