Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event (extended)

USC/ISI Technical Report ISI-TR-2016-709b

May 2016, updated September 2016

Giovane C. M. Moura[1] Ricardo de O. Schmidt[2] John Heidemann[3]
Wouter B. de Vries[2] Moritz Müller[1] Lan Wei[3] Cristian Hesselman[1]
1: SIDN Labs 2: University of Twente 3: USC/Information Sciences Institute

ABSTRACT

Distributed Denial-of-Service (DDoS) attacks continue to be a major threat on the Internet today. DDoS attacks overwhelm target services with requests or other traffic, causing requests from legitimate users to be shut out. A common defense against DDoS is to replicate a service in multiple physical locations/sites. If all sites announce a common prefix, BGP will associate users around the Internet with a nearby site, defining the *catchment* of that site. Anycast defends against DDoS both by increasing aggregate capacity across many sites, and allowing each site's catchment to contain attack traffic, leaving other sites unaffected. IP anycast is widely used by commercial CDNs and for essential infrastructure such as DNS, but there is little evaluation of anycast under stress. This paper provides the *first evaluation of several IP anycast services under stress with public data*. Our subject is the Internet's Root Domain Name Service, made up of 13 independently designed services ("letters", 11 with IP anycast) running at more than 500 sites. Many of these services were stressed by sustained traffic at 100× normal load on Nov. 30 and Dec. 1, 2015. We use public data for most of our analysis to examine how different services respond to stress, and identify two policies: sites may *absorb* attack traffic, containing the damage but reducing service to some users, or they may *withdraw* routes to shift both good and bad traffic to other sites. We study how these deployment policies resulted in different levels of service to different users during the events. We also show evidence of *collateral damage* on other services located near the attacks.

IMC 2016, November 14 - 16, 2016, Santa Monica, CA, USA

 ISBN 978-1-4503-4526-2/16/11…$15.00

DOI: http://dx.doi.org/10.1145/2987443.2987446

1. INTRODUCTION

Although not new, denial-of-service (DoS) attacks are a continued and growing challenge for Internet services [2, 3]. In most DoS attacks the attacker overwhelms a service with large amounts of either bogus traffic or seemingly legitimate requests. Actual legitimate requests are lost due to limits in network or compute resources at the service. Once overwhelmed, the service is susceptible to extortion [41]. Persistent attacks may drive clients to other services. In some cases, attacks last for weeks [17].

Three factors enable today's Distributed DoS (DDoS) attacks: source-address spoofing allows a single machine to masquerade as many machines, making filtering difficult. Second, some protocols amplify attacks sent through a reflector, transforming each byte sent by an attacker into 5 or 500 (or more) bytes delivered to the victim [50]. Third, botnets of thousands of machines are widespread [31], making vast attacks possible even without spoofing and amplification. Large attacks range from 50–540 Gb/s [4] in 2016, and 1 Tb/s attacks are within reach.

Many protocol-level defenses against DNS-based DDoS attacks have been proposed. Source-address validation prevents spoofing [24]. Response-rate limiting [56] reduces the effect of amplification. Protocol changes such as DNS cookies [21] or broader use of TCP [63] can blunt the risks of UDP. While these approaches reduce the effects of a DoS attack, they cannot eliminate it. Moreover, deployment rates of these approaches have been slow [9], in part because there is a

section	observation
§2.2	design choices under stress are withdraw or absorb; best depends on attackers vs. capacity per catchment
§3.1	event was at likely 35 Gb/s (50 Mq/s, an upper bound), resulting in 150 Gb/s reply traffic
§3.2	letters saw minimal to severe loss (1% to 95%)
§3.3	loss was not uniform across each letter's anycast sites; overall loss does not predict user-observed loss at sites
§3.4	some users "flip" to other sites; others stick to sometimes overloaded sites
§3.5	at some sites, some servers suffered disproportionately
§3.6	some collateral damage occurred to co-located services not directly under attack

Table 1: Key observations in this paper.

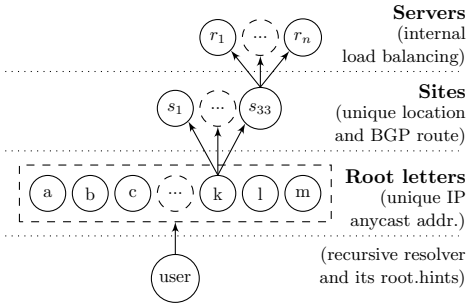


Figure 1: Root DNS structure, terminology, and mechanisms in use at each level.

mismatch of incentives between who must deploy these tools (all ISPs) and the victims of attacks.

Defenses in protocols and filtering are limited, though—ultimately the best defense to a 10,000-node botnet making legitimate-appearing requests is capacity. Services can be replicated to many IP addresses, and each IP address can use IP anycast to operate at multiple locations. Many locations allow a single service to provide large capacity for processing and bandwidth.

Many commercial services promise to defend against DDoS, either by offering DDoS-filtering as a service (as provided by Verizon, NTT, and many others), or by providing a service that adapts to DDoS attacks (such as Akamai [28], Cloudflare, and others). Yet the specific impact of DDoS on real infrastructure has not widely been reported, often because commercial infrastructure is proprietary.

The DNS is a common service, and the root servers are a fundamental, high-profile, and publicly visible service that have been subject to DoS attacks in the past. As a public service, they are monitored [44] and strive to self-report their performance. Perhaps unique among many large services, the Root DNS service is operated by 12 different organizations, with different implementations and infrastructure. Although the internals of each implementation are not public, some details (such as the number of anycast sites) are.

To evaluate the effects of DoS attacks on real-world infrastructure, we analyze two specific events: the Root DNS events of Nov. and Dec. 2015 (see §2.3 for discussion and references). We investigate how the DDoS attack affected reachability and performance of the anycast deployments. This paper is the first to explore the response of real infrastructure across several levels, from specific anycast services (§3.2), physical sites of those services (§3.3), and of individual servers (§3.5). An important consequence of high load on sites is routing changes, as users "flip" from one site to another after a site becomes overloaded (§3.4). Table 1 summarizes our key observations from these studies.

Although we consider only two specific events, we explore their effects on 13 different DNS deployments of varying size and capacity. From the considerable variation in response across these deployments we identify a set of potential responses, first in theory (§2.2) and then in practice (§3). Exploration of additional attacks, and of the interplay of IP anycast and site select at other layers (for example, in Bing [15]) is future work.

The main contribution of this paper is the *first evaluation of several IP anycast services under stress with public data*. Anycast is in wide use and commercial operators have been subject to repeated attacks, some of which have been reported [41, 42, 48, 57, 17, 49, 4], but the details of those attacks are often withheld as proprietary. We demonstrate that in large anycast instances, *site failures* can occur even if the service as a whole continues to operate. Anycast can both *absorb* attack traffic inside sites, and also *withdraw* routes to shift both good and bad traffic to other sites. We explore these policy choices in the context of a real-world attack, and show that *site flips do not necessarily help* when the new site is also overloaded, or when the shift of traffic overloads it. Finally, we show evidence of *collateral damage* (§3.6) on services near the attacks. These results and policies can be used by anycast operators to guide management of their infrastructure. Finally, the challenges we show suggest potential future research in improving routing adaptation under stress and provisioning anycast to tolerate attacks.

2. BACKGROUND AND DATASETS

Before studying anycast services under attack, we first summarize how IP anycast works. We describe the events affecting the Root DNS service on Nov. 30 and Dec. 1, 2015, and the datasets we use to study these events.

2.1 Anycast Background and Terminology

We next briefly review how IP anycast and the Root DNS service works. The Root DNS service is implemented with several mechanisms operating at different levels (Figure 1): a `root.hints` file to bootstrap, multiple IP services, often anycast; BGP routing in each anycast server; and often multiple servers at each site.

letter	operator	sites reported		sites observed
A	Verisign	5	(5, 0)	5
B	USC/ISI	1	(unicast)	1
C	Cogent	8	(8, 0)	8
D	U. Maryland	87	(18, 69)	65
E	NASA	12	(1, 11)	74
F	ISC	59	(5, 54)	52
G	U.S. DoD	6	(6, 0)	6
H	ARL	2	(pri/back)	2
I	Netnod	49	(48, 0)	48
J	Verisign	98	(66, 32)	69
K	RIPE	33	(15, 18)	32
L	ICANN	144	(144, 0)	113
M	WIDE	7	(6, 1)	6

Table 2: The 13 Root Letters, each operating a separate DNS service, with their reported architecture (number of sites with local/global sites [47], B unicast, H primary/backup), plus the count of sites we observe (§3.3).

The Root DNS is implemented by 13 separate DNS services (Table 2), each running on a different IP address, but sharing a common master data source. These are called the 13 *DNS Root Letter Services* (or just the "Root Letters" for short), since each is assigned a letter from A to M and identified as <letter>.root-servers.net. The letters are operated by 12 independent organizations (Verisign operates both A and J), and each letter has a different architecture, an intentional diversity designed to provide robustness. This diversity happens to provide a rich natural environment that allows us to explore how different approaches react to the stress of common attacks.

Most Root Letters are operated using IP anycast [1]. At the time of the analyzed events, only B-Root was unicast [47], and H-Root operated with primary-backup routing [29]. In IP anycast, the same IP address is announced from multiple *anycast sites* (s_1 to s_{33} in Figure 1), each at a different physical location. BGP routing associates clients (recursive resolvers) who chose to use that service with a nearby anycast site. The set of users of each site defines the site's *anycast catchment*.

Larger sites may employ multiple physical servers (r_1 to r_n in Figure 1), each an individual machine that responds to queries. CHAOS queries are a diagnosis mechanism that return an identifier specific to the server [60]. Although their support is optional, responses can be spoofed, and the reply format is not standardized, all letters reply with patterns they disclose or that can be inferred. (Prior studies have confirmed that CHAOS mapping of anycast is generally complete and reliable, validating it against traceroute and other approaches [23].) Properly interpreted CHAOS queries, observed from many vantage points around the Internet (§2.4.1), allow us to map the *catchment* of each anycast site—the footprint of networks that are routed to each sites.

Root Letters have different policies, architectures, and sizes, as shown in Table 2. Some letters constrain routing to some sites to be *local*, using BGP policies (such as `NOPEER` and `NO_EXPORT`) to limit routing to that site to only its immediate or neighboring ASes. Routing for *global* sites, by contrast, is not constrained.

2.2 Anycast vs. DDoS: Design Options

How should an anycast service react to the stress of a DDoS attack? We ground empirical observations (§3) with the following theoretical evaluation of options.

A site under stress, overloaded with incoming traffic, has two options. It can *withdraw* routes to some or all of its neighbors, shrinking its catchment and shifting both legitimate and attack traffic to other anycast sites. Possibly those sites will have greater capacity and service the queries. Alternatively, it can be become a *degraded absorber*, continuing to operate, but with overloaded ingress routers, dropping some incoming legitimate requests due to queue overflow. However, continued operation will also absorb traffic from attackers in its catchment, protecting other anycast sites [1].

These options represent different uses of an anycast deployment. A withdrawal strategy causes anycast to respond as a waterbed, with stress displacing queries from one site to others. The absorption strategy behaves as a conventional mattress, "compressing" under load, with queries getting delayed or dropped. We see both of these behaviors in practice and observe them through site reachability and RTTs.

Although described as strategies and policies, these outcomes are the *result* of several factors: the combination of operator and host ISP routing policy, routing implementations withdrawing under load [54], the nature of the attack, and the locations of the sites and attackers. Some policies are explicit, such as the choice of local-only anycast sites, or operators removing a site for maintenance or modifying routing to manage load. However, under stress, the choices of withdrawal and absorption can also be results that *emerge* from a mix of explicit choices and implementation details, such as BGP timeout values. We speculate that more careful, explicit, and automated management of policies may provide stronger defenses to overload, an area of future work.

Policies in Action: We can illustrate these policies with the following thought experiment. Consider the anycast system in Figure 2, it has three anycast sites: s_1, s_2, S_3, four clients c_0 and c_1 in s_1's catchment, with c_2 in s_2 and c_3 in S_3's. Let A_0 represent both the identity of the attacker and the volume of its attack traffic, and s_1 represent the site and its capacity.

The best choice of defense depends on the relative sizes of attack traffic reaching each site. For simplicity, we can ignore legitimate traffic (c_*), since DNS deployments are greatly overprovisioned ($c_* \ll A_*$). Overprovisioning by 3× peak traffic is expected [14], and 10× to 100× overprovisioning is common. (For example, a

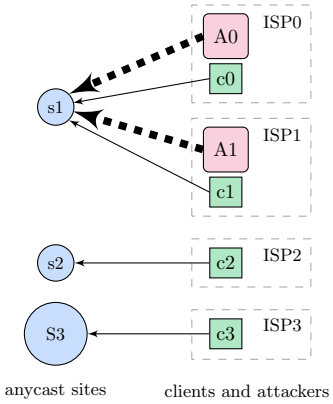


Figure 2: An example anycast deployment under stress.

modest modern computer can handle an entire letter's typical traffic (30–60k queries/s, Table 3), and we see at least 4 to more than 200 servers per letter in our analysis.)

To consider alternative responses to attack we evaluate a deployment where $s_1 = s_2$ and $S_3 = 10s_1$, as attack strength $A_0 = A_1$ increases. We measure the effects of the attack by the total number of served clients (H, "happiness").

1. If $A_0 + A_1 < s_1$, then the attack does not hurt users of the service, $H = 4$.

2. If $A_0 + A_1 > s_1$ and $A_0 < s_1$ (and $A_1 < s_2$), then s_1 is overwhelmed ($H = 2$) but can shed load. If it withdraws its route to ISP1, A_1 and c_1 shift to s_2 and all clients are served: $H = 4$.

3. If $A_0 > s_1$ and $A_0 + A_1 < S_3$, then a attackers can overwhelm a small site, but not the bigger site. *Both* s_1 and s_2 should withdraw all routes and let the large site S_3 handle all traffic, for $H = 4$.

4. If $A_0 > s_1$, $A_0 + A_1 > S_3$, but $A_1 < S_3$, one can re-route ISP1 (with A_1 and c_1) to S_3, for $H = 3$.

5. If $A_0 > S_3$, the attack can overwhelm *any* site; making no change is optimal. s_1 becomes a degraded absorber and protects the other sites from the attack, at the cost of clients c_0 and c_1. $H = 2$.

(Withdrawing routes in response to attacks may also increase latency as catchments change. Our definition of H ignores latency as a secondary factor, focusing only on ability to respond.)

Implications of this model: This model has several important implications, both about the range of *possible* policies, what policies are *practical* today, and directions to explore in the *future*.

This thought experiment shows that for small attacks, the *withdraw* policy can improve service by spreading the attack (although perhaps counter-intuitive, less can be more!). For large attacks, *degraded absorbers* are necessary to protect some clients, at the cost of others. We cannot directly apply these rules in this paper, since we know neither site capacity (something generally kept private by operators as a defensive measure), nor how much attack traffic reaches each site (a function of how attackers align with catchment, again, both unknown to us). Our hope is that the scenarios of this thought experiment can help us interpret our observations of what actually occurs.

A second implication is that choice of optimal strategy is very sensitive to actual conditions—which of the five cases apply depend on attack rate, location, and site capacity. The practical corollary is that choosing the optimal strategy is not easy for operators, either. Attack traffic volumes are unknown to operators, when the attack exceeds capacity; attack locations are unknown, due to source address spoofing; the effects of route changes are difficult to predict, due to unknown attack locations; and route changes are difficult to implement, since routing involves multiple parties. In the face of uncertainty about attack size and location, absorption is a good default policy. However, route withdrawals may occur due to BGP session failure, so both policies may occur.

As an alternative to adjusting routing or absorbing attacks, many websites use commercial anti-DDoS services that do traffic "scrubbing". Such services capture traffic using BGP, filter out the attack, and finally forward the clean traffic to the original destination. While cloud-based scrubbing services have been used by websites (for example, in the 540 Gb/s DDoS attack against the Rio 2016 Olympic Games website [4] or the DoS against ProtonMail [42]), to our knowledge Root DNS providers do not use such services, likely because Root DNS traffic is a very atypical workload (DNS, not HTTP).

Finally, a key implication of this model is that there *can be* better possible strategies than just absorbing attacks. As described above, they require information about attack volume and location that is not available today, but their development is promising future work.

2.3 The Events of Nov. 30 and Dec. 1

On November 30, from 06:50 to 09:30 (UTC), then again on December 1, 2015 from 05:10 to 06:10, many of the Root DNS Letters experienced an unusual high rate of requests [48]. Traffic rates peaked at about 5M queries/s, at least at A-Root [57], more than 100× normal load. We sometimes characterize these events as an "attack" here, since sustained traffic of this volume seems unlikely to be accidental, but the intent of these events is unclear.

An early report by the Root Operators stated that several letters received high rates of queries for 160 minutes on Nov. 30 and 60 minutes on Dec. 1 [48]. Queries used fixed names, but source address were randomized. Some letters saw up to 5 million DNS queries per second, and some sites at some letters were overwhelmed

by this traffic, although several letters were continuously reachable during the attack (either because they had sufficient capacity or were not attacked). There were no known reports of end-user visible errors, because top-level names are extensively cached, and the DNS system is designed to retry and operate in the face of partial failure.

A subsequent report by Verisign, operator of A- and J-Root, provides additional details [57]. They stated that it was limited to IPv4 and UDP packets, and that D-, L-, and M-root were not attacked. They confirm that the event queries used fixed names, with www.336901.com on Nov. 30 and www.916yy.com on Dec. 1. They reported that A and J together saw 895M different IP addresses, strongly suggesting source address spoofing, although the top 200 source addresses accounted for 68% of the queries. They reported that both A- and J-Root were attacked, with A continuing to serve all regular queries throughout, and J suffering a small amount of packet loss. They reported that Response Rate Limiting was effective [57], identifying duplicated queries to drop 60% of the responses, and filtering on the fixed names was also able to reduce outgoing traffic. They suggested the traffic was caused by a botnet.

Motivation: We do not have firm conclusions about the motivation for these events. As Wessels first observed [59], the intent is unclear. The events do not appear to be DNS amplification to affect others since the spoofed sources spread reply traffic widely. They might be a DDoS targeted at services at the fixed names listed above, but .com must resolve those names, not the roots. Also, an attack on the fixed names would be much more effective if the root lookup was cached and not repeated. Possibly it was an attack on those targets that went awry due to bugs in the attack code. It may be a direct attack on the Root DNS, or even a diversion of other activity. Fortunately, the intent of the event is irrelevant to our use of the event to understand anycast systems under stress.

Generalizing: We analyze and provide data for both events. Subsequent root events [49] differ in the details of the event, but pose the same operational choices of how to react to an attack (§2.2).

We focus on specific IP anycast services providing DNS under stress. Root DNS is provided by multiple such services, and CDNs add DNS-based redirection as another level of redundancy [15]. Although we briefly discuss overall performance (§3.2.2), full exploration of these topics is future work that can build on our analysis of IP anycast.

2.4 Datasets

We use these large events to assess anycast operation under stress. Our evaluation uses publicly available datasets provided by RIPE, several of the Root operators, and the BGPmon project. We thank these organizations for making this data available to us and other researchers. We next describe these data sources and how we analyze it. The resulting dataset from the processing described next is publicly available at our websites [40].

2.4.1 *RIPE Atlas Datasets*

RIPE Atlas is a measurement platform with more than 9000 global devices (Atlas Probes) that provide *vantage points* (*VPs*) that conduct network measurements [30, 46]. All Atlas VPs regularly probe all Root DNS Letters. A subset of this data appears in RIPE's DNSMON dashboard evaluating Root DNS [44]. RIPE identifies data from all VPs that probe each root letter with a distinct measurement ID [45]. Our study considers all available Atlas data (more than DNSMON reports), with new processing as we describe below.

RIPE's baseline measurements send a DNS CHAOS query to each Root Letter every 4 minutes. At the time of the event, A-Root was an exception and was probed only every 30 minutes, too infrequent for our analysis (§3.2) (it is now probed as frequently as the other letters). Responses to CHAOS queries are specific to root letters (after cleaning, described below) but each letter follows a pattern that can be parsed to determine the site and server that VP sees. For this report we normalize identification of roots in the format *X-APT*, where *X* is the Root Letter (A to M) and *APT* is a three-letter airport code near the site.

Due to space limitations, we provide examples of specific letters rather than reporting data for all anycast deployments. We focus predominantly on E- and K-Root, since they provide anycast deployments with dozens of sites. These examples concretely illustrate of the operational choices (§2.2) all anycast deployments face.

Data cleaning: We take several steps to clean RIPE data for using it in our analysis. Cleaning preserves nearly all VPs (more than 9000 of the 9363 currently active in May 2016), but discards data that appears incorrect or provides outliers. We discard data from VPs with Atlas firmware before version 4570. Atlas firmware is regularly updated [43], and version 4570 was released in early 2013. Out of caution, we discard measurements from earlier firmware on non-updating VPs to provide consistent (current) methods of measurement. Moreover, we also discard measurements of a few VPs where traffic to a root appears to be served by third parties. We identify hijacking in 74 VPs (less than 1%) by the combination of a CHAOS reply that does not match that letter's known patterns and unusually short RTTs (less than 7 ms), following prior work [23].

After cleaning we map all observations into a time series with ten-minute bins. In each time bin we identify, for each Root Letter, the response: either a site the VP sees, a response error code [39], or an absence of a reply after 5 seconds (the Atlas timeout). Each time bin represents 2.5 RIPE probing intervals, allowing us to synchronize RIPE measurements that otherwise occur at arbitrary phases. (When we have differing replies

in one bin, we prefer sites over errors, and errors over missing replies.)

Limitations of RIPE Atlas: RIPE Atlas has known limitations: although VPs are global, their locations are heavily biased towards Europe. This bias means Europe is strongly over-represented in per-letter reachability (§3.2), but it does not influence our analysis of specific user behavior (§3.4). The largest risk uneven distribution of VPs poses is that some anycast sites may have too few VPs to provide reliable reporting. While we report on all anycast sites we observe, we only consider sites whose catchments contain a median of at least 20 VPs during the two days.

In addition, RIPE VPs query specific Root letters, so they do not represent "user" queries. (Regular user queries employ a recursive resolver selects one or more letters to query.) We take advantage of this approach to study specific letters and sites (§3), but it prevents us from studying Root DNS reachability as a whole (§3.2.2).

Finally, VPs fail independently. We focus our attention on sites typically seen by 20 or more VPs to avoid bias from individual VP failure over the two days.

2.4.2 *RSSAC-002*

RSSAC-002 is a specification for operationally-relevant data about the Root DNS [51]. It provides daily, per-letter query rates and distributions of query sizes.

All Root Letters have committed to provide RSSAC-002 data by 2017. At the time of the events, only five services (A, H, J, K, and L) were providing this data [47]. In addition, RSSAC-002 monitoring is a "best effort" activity that is not considered as essential as operational service, so reporting may be incomplete, particularly at times of stress.

2.4.3 *BGPmon*

We use BGP routing data from BGPmon [61]. BGPmon has peers to dozens of routers providing full routing tables from different locations around the Internet. We use data from all available peers on the event days (152 peers) to evaluate route changes at anycast sites in §3.4.1.

3. ANALYSIS OF THE EVENTS

To evaluate the events we begin with overall estimates of their size, then drill down on how the events affected specific Root Letters, sites in some letters, and individual servers at those sites. We then reconsider the effects of the attack as a whole, both on Root DNS service and on other services.

3.1 How Big Were the Events?

We next estimate the size of the events. Understanding the size is important to gauge the level of resources available to the traffic originator. We begin with RSSAC-002 reports, but on Nov. 30, only a few letters provided this data, and as previously described (§2.4.2), best-effort RSSAC-002 data is incomplete. We therefore estimate an upper-bound on the event based on inference from available data.

RSSAC-002 statistics over each day, so to estimate the event size we define a baseline as the mean of the seven days before the event. We then look at what changed on the two event days (A-Root had an independent attack on 2015-11-28, so we drop this data point and scale proportionally). Query sizes are reported in bins of 16 bytes. Verisign stated that the attacks were of specific query names (see §2.3), and RSAAC-002 reports query sizes in bins of 16 bytes, allowing us to identify attacks by unusually popular bins. For queries, the 32-to-47 B bin on Nov. 30 and the 16-to-32 B bin on Dec. 1 while response sizes were between 480 and 495 bytes for both events. These sizes are for DNS payload only. We confirm total traffic size (with headers) in two ways, both by adding 40 bytes to account for IP, UDP, and DNS headers, and by generating queries with the given attack names. We confirm full packets (payload and header) of 84 and 85 bytes for queries and 493 or 494 bytes for responses, consistent with RSSAC-002 reports. We use these sizes to estimate incoming bitrates.

Table 3 gives our estimates on event traffic from the five letters reporting RSSAC-002 statistics. The baseline (right column) is only 1–10% of attack traffic (mean: 3%); we subtract the baseline from queries and responses therefore our estimations show the only the *extra* (Δ) traffic caused by the events. These reported values differ greatly across letters and between queries and responses. We believe differences across letter represent measurement errors, with most letters under-measuring traffic when under attack (under-reporting is consistent with large amounts of lost queries described in §3.2). We see fewer responses than requests, likely because of Response Rate Limiting [56] which suppresses duplicate queries from the same source address [59]. We provide both a lower-bound on attacks that considers only known event traffic, and a scaled value that accounts for the six sites known to have been attacked that did not provide RSSAC-002 data at event time. This lower bound has a large underestimate because 3 of the 4 reports were known to drop event traffic, and there is an approximate 3% overestimate by including baseline queries.

We propose an upper-bound for event size by correcting for both of these types of under-reporting. To correct, we accept that A-Root's RSSAC-002 data measured the entire event. Verisign reported [59] A-Root graphs of input traffic showing about 5Mq/s at both A- and J-Root (although J's RSSAC-002 reports are much lower). They also report that 10 of 13 letters were attacked (D, L, and M were not attacked). We add the assumption that all attacked letters received equal traffic. We confirm this assumption in two ways. First, B-Root can confirm [5] that it saw offered load around

RSSAC reports	2015-11-30 (160 min.) Δ queries Mq/s	Gb/s	M IPs (ratio)	Δ responses Mq/s	Gb/s	2015-12-01 (60 min.) Δ queries Mq/s	Gb/s	M IPs (ratio)	Δ responses Mq/s	Gb/s	*Baseline* queries Mq/s	M IPs
A	5.12	3.44	1,813 (340×)	3.84	15.13	5.21	3.54	1,345 (253×)	3.93	15.53	*0.04*	*5.35*
H	0.23	0.15	36.14 (13.3×)	—	—	0.32	0.22	16.22 (6.5×)	—	—	*0.03*	*2.94*
J	1.90	1.28	765.24 (280×)	1.10	4.32	2.29	1.56	355.68 (129×)	1.43	5.66	*0.05*	*2.78*
K	1.07	0.72	39.23 (14.4×)	0.48	0.32	1.12	0.76	40.88 (15.0×)	0.28	1.09	*0.04*	*2.92*
L	0.05*	0.04*	36.15 (13.3×)*	0.05*	0.19*	0.10*	0.07*	16.22 (6.5×)*	0.09*	0.37*	*0.06*	*2.94*
bounds (lower and **upper**):												
lower	8.32	5.59	–	5.42	19.77	8.94	6.08	–	5.64	22.28	0.22	–
(scaled)	(20.8)	(14.0)	–	(13.5)	(49.4)	(22.4)	(15.2)	–	(14.1)	(55.2)	–	–
upper	**51.22**	**34.42**	–	**38.37**	**151.31**	**52.09**	**35.42**	–	**39.31**	**155.35**		

Table 3: RSSAC-002 reports for daily IPv4/UDP traffic for the two days of events, subtracted from the a 7-day mean baseline, and lower- and upper-bounds on event sizes. *L-Root was not attacked and therefore excluded from lower and upper bounds.

5 Mq/s, consistent with A-Root's statement. Second, we can infer event sizes by comparing accepted traffic loads in Table 3 with observed loss rates from Figure 3. That suggests H-Root should have received 1.6 Mq/s, J-Root about 2.44 Mq/s, and K-Root about 1.6 Mq/s.

Given these reports and this assumption, our best estimate of actual attack strength is somewhere between half and all of our upper-bound estimate. This estimate is somewhat rough, but it provides strong evidence that this was not a small attack. While 6× more than the lower-bound of directly observed traffic, this estimate reflects significant query loss that occurs during the event and in measurement systems tuned for regular operation.

If our upper-bound estimate is correct, the aggregate size of this attack across all letters is about 35–40 Gb/s. Although attacks exceeding 100 Gb/s have been demonstrated since 2012 [2, 4], such large attacks are usually performed using amplification [50] (for example, as the reply traffic of 151 Gb/s on Table 3). Directly sourced traffic of 35 Gb/s on the roots therefore represents a large attack.

We can also see for all letters a large increase (by a factor 6.5× to 340×) in the number of unique IPv4 addresses observed by each letter during the attacks. This observation conforms with the initial reports on the use of IP spoofing during these attacks [59].

3.2 How Were Individual Letters Affected?

We next consider how each letter reacted to the event and measure overall Root DNS performance. Letter-specific queries from RIPE Atlas show that individual root letters suffered minimal to severe loss rates. We caution that these loss rates do not directly translate to end-user delays, since recursive resolvers cache and retry against different letters (§3.2.2), and since users interact with specific anycast sites (§3.3).

3.2.1 Reachability of Specific Letters

Figure 3 shows the reachability for each Root Letter from RIPE Atlas. We plot D-, L-, and M-Root together because they see no visible change, consistent with reports that they were not attacked [57]. (In §3.6 we later show that a few D-root sites appear slightly affected by the event.) On these dates, Atlas probed A-Root less frequently than other letters (§2.4.1), so in this graph we scale A's observations to account for this difference. Because infrequent probing of A-Root makes the event dynamics impossible to discern, we omit A-Root from analysis in the rest of the paper.

All the other letters experience different degrees of reachability problems during the reported attack intervals (§2.3). There is a strong correlation ($R^2 = 0.87$) between how many sites a letter has (Table 2) and their worst responsiveness, measured by the smallest number of Atlas VPs that successfully receive responses during the events (more sites→more VPs receive responses). B-Root, a unicast letter, suffered the most, followed by H, with two sites and primary-secondary routing. With many sites, J-Root sees some VPs lose service, but only a few. We evaluate the *causes* for service loss in §3.3, but this correlation reflects some combination of more sites providing greater aggregate capacity and isolating some users from some event traffic.

We can also evaluate overall performance for each letter by the RTT of successful queries, as shown in Figure 4. Note that each letter has a different baseline RTT, corresponding with the median distance from Atlas VPs to anycast sites for that letter. Although B-Root suffered the most in terms of reachability Figure 3, it experienced little change in RTT when queries *were* successful. G- and H-Root, in turn, see large changes in latency. In the next section we show that anycast sites can *fail*, causing routing to shift their traffic to other locations. Thus we believe these shifts in RTT indi-

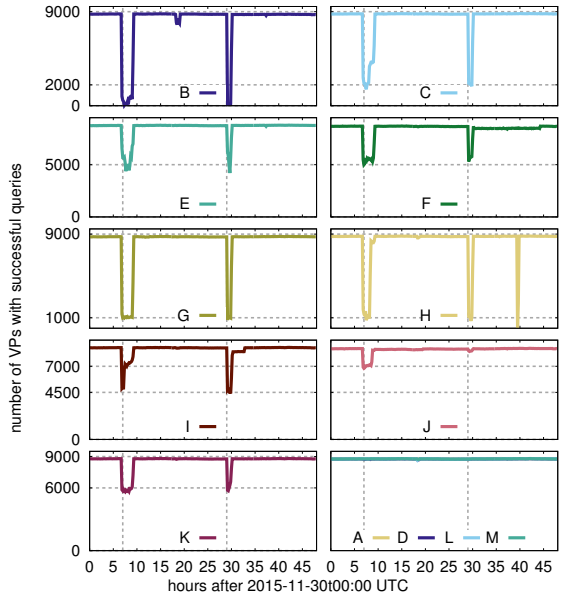


Figure 3: Number of VPs with successful queries (in 10-minute bins). (All plots are scaled consistently, with nearly 9000 VPs across 48 hours of observation. In all graphs, dotted lines highlight approximate event start times. Here they also show the lowest values for the dips.)

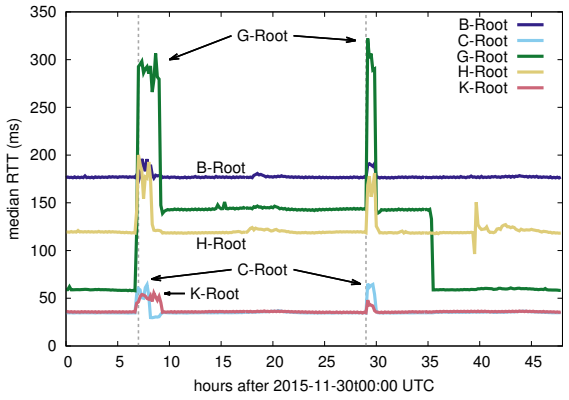


Figure 4: Median RTT for some letters during the attacks. Letters with no significant change (A, D, E, F, I, J, L, and M) are omitted.

cate route changes that shift VP traffic to more distant sites. For example, H-Root has sites on the U.S. East and West coasts (north of Baltimore, Maryland, and in San Diego, California). Most Atlas VPs are in Europe, so we infer that the primary site for H is the U.S. East coast, but when that route is withdrawn (during both events) traffic shifts to the west coast. This assumption is confirmed by H's median RTT at that time matching B-Root's RTT, since B-Root is also on the U.S. West coast. We examine site route withdrawals in more details in §3.3.

3.2.2 *Reachability of the Root DNS as a Whole*

While we see that individual letters show degraded responsiveness under stress, the DNS protocol has several levels of redundancy, and a non-response from one letter should be met by a retry at another letter. This paper does not evaluate overall responsiveness of the Root DNS, but our per-letter analysis shows some evidence of this redundancy.

L-Root was not subject to these attacks [59], yet Table 3 it exhibited that L-Root shows a significant increase in query rate during the second event, with a 1.66× increase in queries-per-second. More impressively, it sees a 6- or 13-fold increase in number of unique IPs on both event dates. We later describe "site flips", where VPs change anycast sites (§3.4.1); this coarse data suggests *letter flips* also occur, as recursive resolvers switch from one letter to another, perhaps to prefer a shorter RTT [62, 36]. While not the focus of this paper, these letter flips show the multiple levels of resilience in the Root DNS system.

3.3 How Were Anycast Sites Affected?

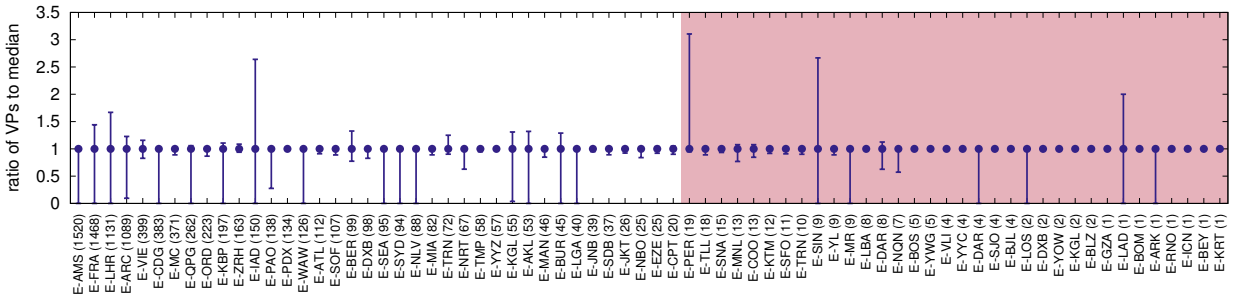
Overall loss rates for each letter (§3.2) may suggest that query loss is uniform for all who use that letter. We next show that these loss rates are *not* uniformly seen by all users. Anycast services are composed of multiple sites (Table 2), and anycast operators and their hosting ISPs can design sites to withdraw routes or continue as degraded absorbers when under stress (§2.2). We next look at behavior across all sites of a given letter to identify evidence of these policies in action.

3.3.1 *Site Reachability*

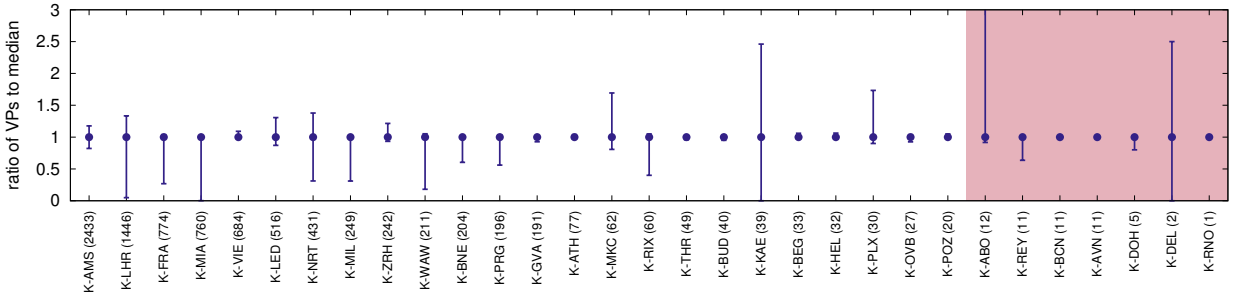
We first consider site reachability: how many Atlas VPs reach a letter's sites over the two days of observations, measured in each ten-minute bin. The median number of VPs over the observation provides a baseline of "regular" behavior, calibrating how RIPE Atlas maps to a given service. Atlas coverage is incomplete; some sites have zero or a few VPs, while others have thousands in their regular catchment. Our use of median normalizes coverage to identify trends, such as if the site adds or loses VPs. Addition of VPs to one site indicates withdrawal of routes to another site, possibly in reaction to stress. Reduction in VPs indicates that either that site withdrew some or all routes, or that it was overloaded and simply lost queries—reduction can therefore be caused by both withdrawal and absorption.

Figure 5 shows all sites for two letters (E- and K-Root, selected as representatives with many sites). Numbers in parenthesis show the median number of VPs at each site, while the lines show how much that site shrank or grew over the two days, normalized to the median.

We see that sites show two responses indicating reduced capacity. Some (such as E-AMS) become completely unavailable, as shown by the minimum drop-



(a) E-Root sites



(b) K-Root sites

Figure 5: Minimum and maximum number of VPs, normalized to median (shown in parenthesis per individual site), for sites from E- and K-Root. Observations are grouped into 10-minute bins over two days. Sites are ordered by median number of VPs, and the red, shaded area highlights sites with fewer than 20 VPs (our threshold for stability, §2.4.1)).

ping to zero; some become nearly unavailable, such as K-LHR; K-Root confirmed unavailability of some sites [64]. Others (E-NRT, K-WAW) become partially available.

In addition, several sites show an increase above median over the period (the maximum blue value is greater than 1). Several of the well-observed K-Root sites show some increase (K-AMS, K-LHR, K-LED, K-NRT), as do many of the well-observe E-Root sites (E-FRA, E-LHR, E-ARC, E-VIE, E-IAD).

We confirm that these swings in catchments are directly correlated with the events and are not typical behavior. We repeated the analysis of Figure 5 over two days during the week following the events (2016-12-05 and 2016-12-06)(See Figure 16 in Appendix A.) . On these "normal" days, considering sites with reasonable visibility (20 or more VPs, so medians are stable), we see *no* variation in VPs per site for K-Root, and only minor variation (mostly within 8%) for 13 sites of E-Root.

Second, Figure 6 shows the size of each site's catchment during the events, for E- and K-Root. Each miniplot represents one site, with the line showing how many VPs are mapped to it relative to the site's median. From this figure we see that sites from these two letters behaved completely different. While most sites of E-Root either see an increase or a decrease on their reachability, most sites of K-Root seem to overlook the attack. (Note that large increases observed for few sites, such as E-DXB and K-DEL, are caused by a very low median (two VPs)—any additional VP hitting this sites during the attack can cause a peak on reachability.)

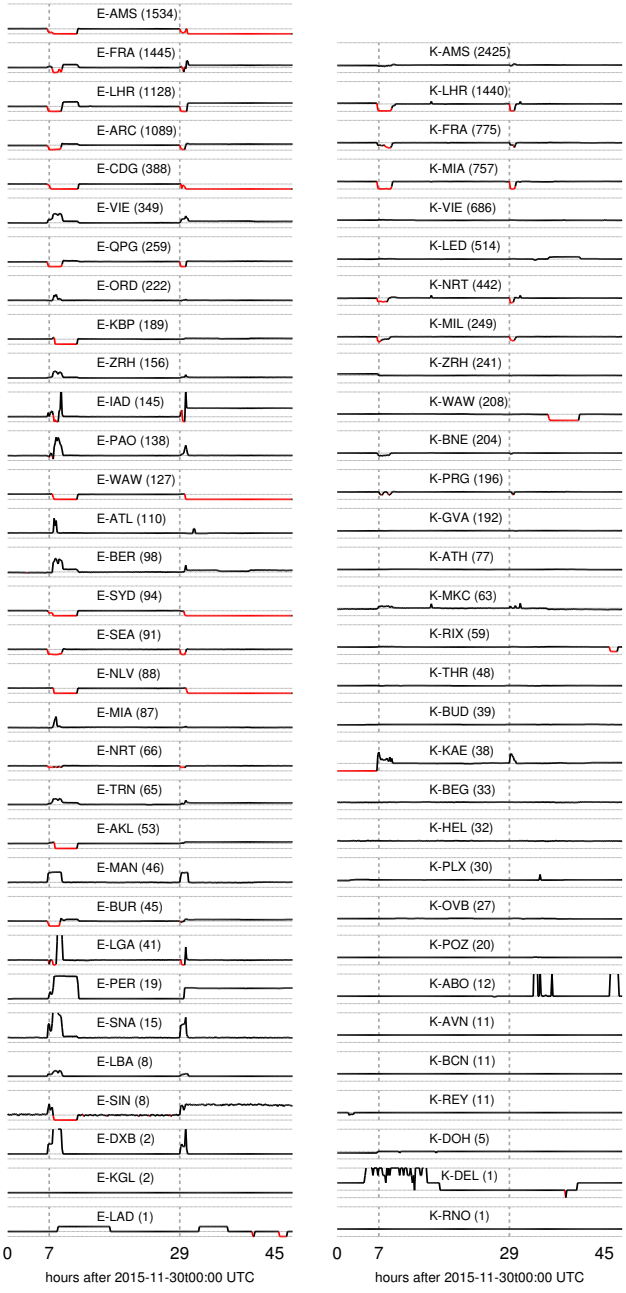
Figure 6 shows that five sites from E Root (E-AMS, E-CDG, E-WAW, E-SYD and E-NLV) seem to "shut down" after the attack of Dec. 1 (hour 29). These sites also had reachability strongly compromised during the first event on Nov. 30 (hour 7).

What is interesting to see for the sites of both letters in Figure 6 is that sites with large numbers of median VPs in their catchments showed reachability problems. An exception is K-AMS, with a large number of VPs in its catchment, which took on more traffic than usual during the whole period.

For E-Root, sites that show an increase over median suggest that some other sites are withdrawing some routes at other sites. However, that does not explain why letters show reduced overall reachability (Figure 3): if overloaded sites fail and traffic shifts, all queries should be answered. We next look for evidence of degraded absorption.

3.3.2 *Site RTT Performance*

To assess if sites that remain accessible are overloaded (implying they operate as degraded absorbers), we next examine RTT of successful queries.



(a) E-Root sites (b) K-Root sites

Figure 6: Reachability seen by VPs that received positive responses (RCODE 0) for sites of E- and K-Root. The central line in each plot is the median, with the lower line 0 and the upper line 5× and 3× the median for E- and K-Root respectively. Red lines below the median indicate potential *critical* moments in which reachability dropped below the by the median number of VPs that can normally reach the site. Sites are sorted by median, in parenthesis.

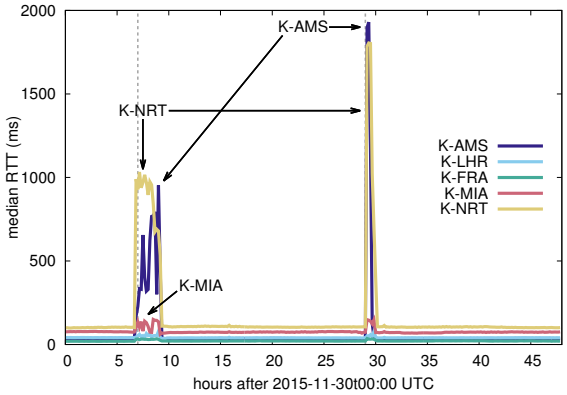


Figure 7: Performance for selected K-Root sites.

Figure 7 shows the median RTT for some K-Root sites that show stress during the events. Although the K-AMS site remained up and showed minimal loss, its median RTT showed a huge increase: from roughly 30 ms to 1 s on Nov. 30, and to almost 2 s on Dec. 1, strongly suggesting the site was overloaded. K-NRT shows similar behavior, with its median RTT rising from 80 ms to 1 s and 1.7 s in the two events. Overload does not always result in large latencies. B-Root (a single site) showed only modest RTT increases (Figure 4), since only few probes could reach it during the attack (Figure 3). We hypothesize that large RTT increases in sites performance are the result of an overloaded link combined with large buffering at routers (industrial-scale bufferbloat [27]).

3.4 How Can Services Partially Fail?

We have shown that letters report different amounts of service degradation (Figure 3), and that their sites seem to follow two policies under stress (§3.3). We next look at service reachability from a *client* perspective to understand how services can partially fail, and how some clients see persistent failures.

3.4.1 *Site Flips: Evidence of Stress*

A design goal of DNS and IP anycast is that service is provided by multiple IP addresses (DNS) and sites (anycast). Through their recursive resolvers, clients can turn to service on other IP address (other Root Letters), and through route changes at upstream ISPs, to other anycast sites. A recursive DNS resolver will automatically retry with another name server if the first does not respond, which is intentional redundancy in the protocol and an operational best practice [22, 62]. Redundancy *inside* most letters depends on IP anycast, and the routing policies DNS service operators establish at each anycast site (withdraw or absorbing, as in §2.2).

To study a client's view of IP anycast redundancy, we look for changes in site catchments. We measure these as *site flips*: when a VP changes from its current anycast site to another. We expect each VP to have a

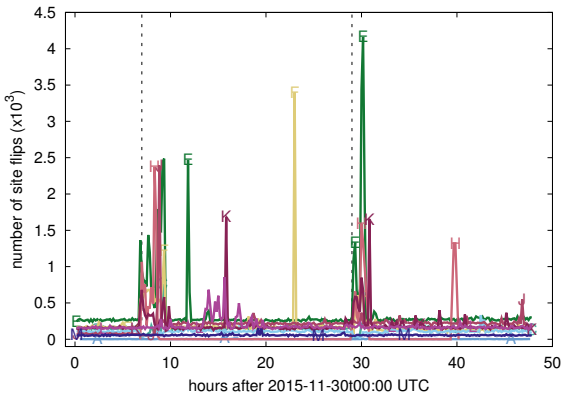


Figure 8: Number of site flips per Root letter.

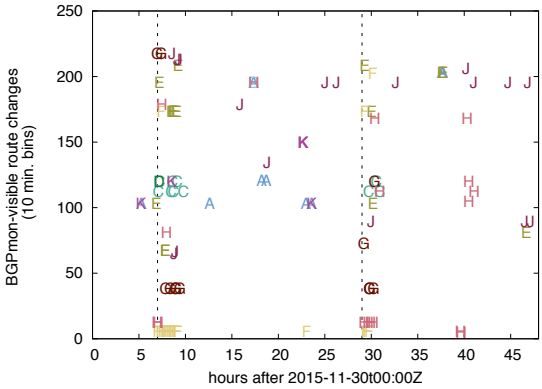


Figure 9: Route changes for each Root Letter (10 minute bins, seen from BGPmon route collectors).

preferred site (hopefully with low RTT), and site flips to be rare, due to routing changes or site maintenance.

Figure 8 shows site flips measured in RIPE Atlas VPs, with bursts of site flips during the event periods for letters that saw event traffic. All letters see thousands of site flips during the event (note the scale of the y-axis), with E, H and K seeing many flips while C, I and J see fewer.

To evaluate if these site flips are actually due to route withdrawals, we use route data from BGPmon (§2.4.3). These BGPmon VPs are in different locations from our RIPE Atlas VPs, so we do not expect them to see exactly the same results, but, if there were route withdrawals, we expect to see more routing activity during the events.

Figure 9 shows the route changes we observe across all Root Letters. With BGPmon VPs and Root anycast sites around the world, we see occasional route changes over the whole time period. With 152 VPs, a routing change near one site can often be seen at 100 or more VPs. But the *very frequent* sets of changes shown by *many* letters in the two event periods (4 to 6 hours and around 29 hours) suggests event-driven route changes

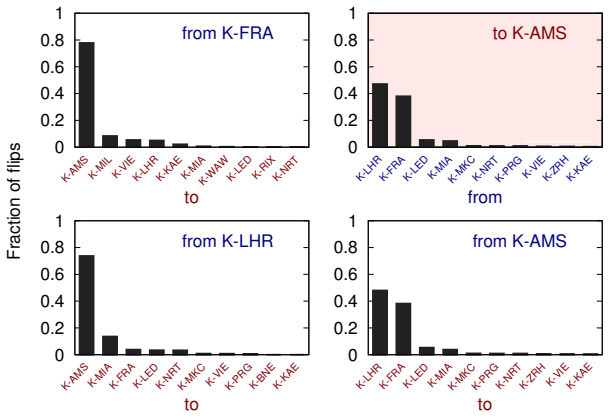


Figure 10: Site flips for selected K-Root instances over the two days.

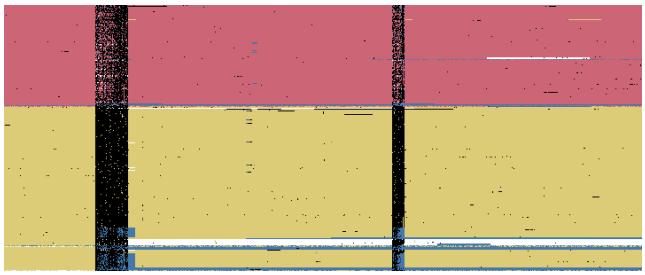
for many letters (C, E, F, G, H, J, K). Route changes for K-Root do not appear at our BGP observers for the second event, and K's BGP changes are lower than we expect based on site flips. We suspect that our BGP vantage points are U.S.-based, while site flips are VPs that are much more numerous in Europe.

3.4.2 *Case Study: K-Root*

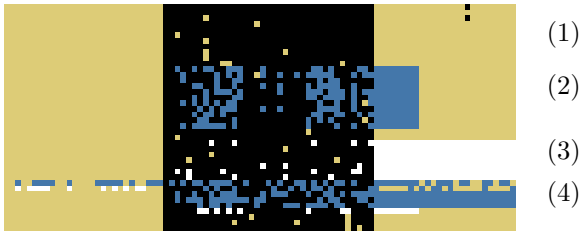
We next consider K-Root as a case study to show what site flips mean in practice. K-Root's sites provide good examples of different policies under stress. We next consider VPs (one per row) that start at K-LHR and K-FRA (London and Frankfurt) to see what happened to these clients during the event. We select these sites to illustrate possible design choices (§2.2) and because they lost nearly all or about half of the VPs during the event; they were more strongly affected then most K-Root sites. From Figure 3 we know that some clients were unsuccessful, while the maximums in Figure 5b show that some sites gained clients.

Figure 10 shows where sites from K-LHR and K-FRA went over the measurement period—the left two graphs show that about 70-80% of all VPs that shifted traffic during the events shift to K-AMS (Amsterdam). The top right (red background) graph shows where new VPs that see K-AMS just were, confirming they mostly arrive from K-LHR and K-FRA. The bottom right graph shows that K-AMS sites also shift back to K-LHR and K-FRA as their preferred catchments after the events.

However, we still ask: *if traffic shifts to other sites and K has excess capacity, why do some VPs fail to reach K during the attack?* VP query failure must result from routing policies and implementation details (§2.2) at each site and its hosts: those policies and details can result in a site that will continue to receive traffic from its peer and operate as a degraded absorber, or that will withdraw its route and reallocate its catchment. We see evidence of both outcomes.



(a) A sample of 300 VPs; start 2015-11-30t00:00Z for 36 hours.



(b) A smaller sample: 40 K-LHR-preferring VPs around the first event.

Figure 11: A sample of 300 VPs for K-Root that start at K-LHR (yellow or light gray) and K-FRA (salmon or medium gray), with locations before, during, and after attacks. Other sites are K-AMS (blue or dark gray), with white indicating other K sites, and black fail on getting a response (timeout). Dataset: RIPE Atlas.

To demonstrate these policies at work, we must look at the actions of individual VPs. Figure 11 shows 300 randomly selected VPs that start at K-LHR (yellow or light gray) and K-FRA (salmon or medium gray) for 36 hours. Each pixel represents the site choice of that VP in 4-minute bins. Black indicates the VP got no reply, while blue (or dark gray) and white indicate selection of K-AMS or some other K-Root site.

We focus on the 40 VPs shown in Figure 11b and see two behaviors during the event and three after. During the event, the top 10 VPs (labeled (1)) stick to K-LHR, but only get occasional replies. They represent a degraded absorbing peering relationship; these clients seem "stuck" to the K-LHR site. The next group labeled (2) shift to K-FRA (salmon or medium gray) during the event and for a short period after, then return to K-LHR. However, during their visit to K-AMS only about a third of their queries are successful. This group shows that K-AMS is overloaded but up, and that these VPs are in ASes that are not bound to K-LHR. For the third group, marked (3), some stay at K-LHR during the event, while others shift to other sites, but all find other sites after the event. Finally, the group (4) shifts to K-FRA during the event and remains there afterward. We see similar groups for the K-FRA sites in the first event and for both sites in the second event.

We believe this kind of *partial failure* represents a *success* of anycast in isolating some traffic to keep other sites functional, but this degraded absorbing policy results in some users suffering during the event due to the overload at K-LHR. While this policy successfully protects most K-Root sites during the event, it also suggests opportunities for alternate policies during attack. Rather than let sites fail or succeed, services may choose to control routing to engineer traffic to provide good service to more users. Alternatively, if attack traffic is localized, services may choose to target routing so that only one catchment is affected—a policy particularly appropriate for attacks where all traffic originates from a single location, even if it spoofs source addresses.

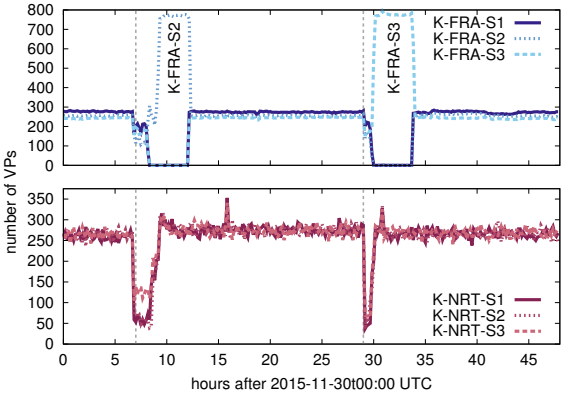


Figure 12: Reachability for individual servers from K-FRA (top) and K-NRT (bottom).

3.5 How Were Individual Servers Affected?

Large anycast sites may operate multiple servers behind a load balancer (Figure 1). We now examine how the events affected individual servers within specific anycast sites. We look at two sites of K-Root, K-FRA and K-NRT as examples, selected because they show different responses to stress. These behaviors are also seen at other sites, but we do not identify or count behaviors across all sites. These examples show it is important to use measurement strategies that consider all servers at a given site.

Figure 12 shows a time series of servers that respond at K-FRA (top) and K-NRT (bottom) during the events. At K-FRA, we typically saw replies from each of the three servers. As the load of each event rose, replies shifted to come from only one server, with none from the other two we previously saw replying. Which server responded was different in the two events, with K-FRA-S2 replying in the first event and -S3 in the second. We do not know if the other two servers failed, or if they were only serving attack traffic, or if traffic from these

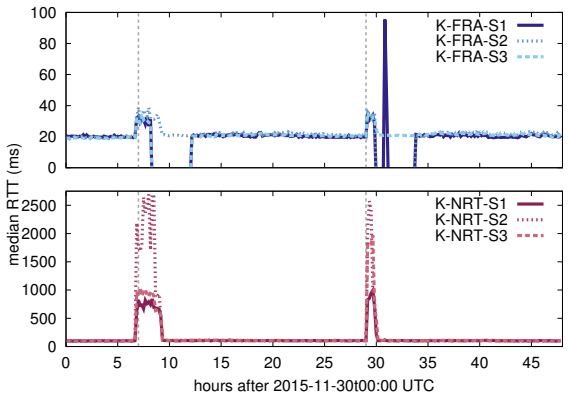


Figure 13: Performance for individual servers from K-FRA (top) and K-NRT (bottom).

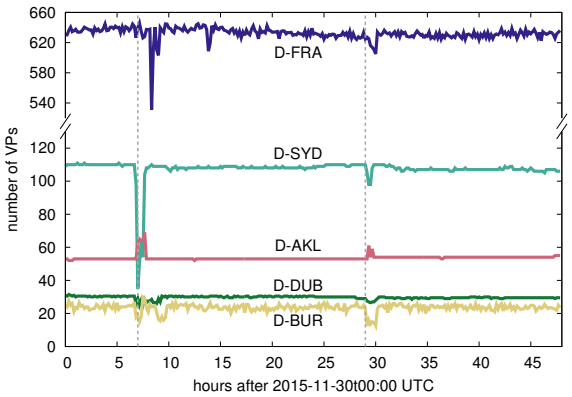


Figure 14: Reachability of those D-Root sites that were affected by the DDoS.

VPs was somehow isolated from attack traffic. Either way, this strategy seems to work reasonably well since Figure 13 shows that, after a short increase in RTT at the beginning of the attack, the median RTT for K-FRA remains stable for successful replies throughout the attack. However, K-FRA seems to be overloaded and dropping queries, as shown in Figure 6b and Figure 11b.

K-Root's Tokyo site (K-NRT) shows a different result. Figure 12 (bottom) shows that VPs had difficulty reaching all three servers from K-NRT during the events. This difficulty suggests that the events affected all K-NRT servers, either because load balancing was mixing our observations with attack traffic, or because attack traffic was congesting a shared link. Figure 13 (bottom) shows larger latencies for successful queries at K-NRT, perhaps suggesting queuing at the router. We also observe that K-NRT-S2 seems more heavily loaded than the other two servers at K-NRT.

These examples show that individual *server* performance and reachability may not reflect overall *site-wide* performance and reachability. Measurement studies of anycast services should therefore insure they study all servers at a site (not just specific servers) to get a complete picture of site and end-user-perceived performance.

3.6 Are There Signs of Collateral Damage?

Servers today, such as the Root DNS servers we study, sometimes are located in data centers that are shared with other services. These services may be unrelated, other infrastructure (such as other top-level domains, TLDs), or even other Root DNS sites. Co-locating services creates some degree of shared risk, in that stress on one service may spill over into another causing *collateral damage*. Collateral damage is a common side-effect of DDoS, and data centers and operators strive to minimize collateral damage through redundancy, overcapacity, and isolation. Prior reports describe it as a problem but provide few details [42].

Hosting details are usually considered proprietary, and commonality can exist at many layers, from the physical facility to peering to upstream providers, making it difficult to assess shared risk. From public data we therefore cannot establish direct causation in a specific common point. Instead, we assess shared risk by *end-to-end evaluation*: we look for service problems in other services not directly the target of event traffic. We study two services: D-Root, a letter that was not directly attacked [57], and the `.nl` TLD. They are chosen because they both show reduced end-to-end performance with timing consistent with the events, strongly suggesting a shared resource with event targets.

D-Root: Figure 14 shows the absolute counts of the number of RIPE Atlas VPs that reach several D-Root sites. D-Root has many sites; we report only subsets that had at least a 10% decrease in reachability during the time of the attacks and were reached by at least 20 RIPE Atlas probes.

These figures show that D-FRA and D-SYD sites both lost VPs during the event. Which data centers host these sites is not public, but correlation of these changes with the events suggests potential collateral damage. (Recall that RIPE VPs probe only one letter, so a reduction in VPs to one site implies either query loss or re-routing, not switching to another letter.)

Frankfurt: There are seven Root Letters hosted in Frankfurt (A, C, D, E, F, I, and K), and we previously observed that traffic shifted to K-FRA and yet that site suffered loss (§3.4.2).

D-FRA sees only small decreases in traffic, suggesting it was only slightly affected by the events to sites for other letter in the same city. However, this change indicates some collateral damage for D-FRA.

The .nl Top-Level Domain: Finally, we have also observed collateral damage at servers that are not part of the Root DNS. We see evidence for collateral damage occurring to the `.nl` top-level domain. In addi-

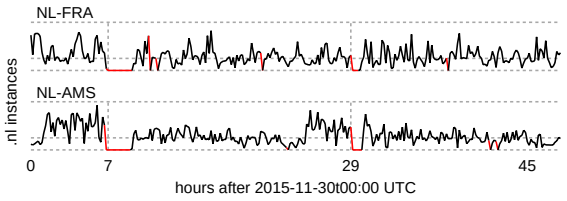


Figure 15: Normalized number of queries for `.nl`, measured at the servers in 10 minutes bins.

tion to four unicast deployments, SIDN operates `.nl` on multiple anycast services. Figure 15 shows query rates for two anycast deployments located near Root DNS servers (exact rates and locations are anonymized). We see both sites show nearly no queries during both events. As a result of this collateral damage, during this period, `.nl` service was carried by other `.nl` servers.

4. RELATED WORK

Distributed Denial-of-Service attacks is a broad area of study and it has been addressed from many different angles in the past years. Studies have shown that DDoS attacks are effective [58].

DDoS attacks are common and growing: Arbor has documented their increasing use and growth in size [2, 3], and there has been DDoS attacks currently reaching 540 Gb/s [4]. Very large attacks often use different protocols to amplify basic attack traffic [50, 55, 20]. Yet DDoS-for-hire ("Booter" services) are easily available for purchase on the gray market—for only a few U.S. dollars, Gb/s attacks can be ordered on demand [52, 32].

Some approaches have been proposed to mitigate amplification [56, 33], spoofing [24], or collateral damage [18]. The continued and growing attacks show that mitigation has been incomplete and that spoofing remains widespread [9].

Many studies have looked at the Root DNS server system, considering performance [11, 26, 12, 35, 19, 7, 53, 38, 16, 34, 23, 37, 8], client-server affinity [53, 10], and effects of routing on anycast [6, 13], as well a proposal to improve anycast performance in CDNs [25]. We draw on prior measurement approaches, particularly the use of CHAOS queries to identify anycast catchments [23].

Closest to our work are prior analyses of the Nov. 30 events [48, 59, 57, 64]. These reports lend insight into the events, but were high level [48, 64] or reported only on specific letters [59, 57, 64].

To the best of our knowledge, our paper is the first to combine multiple sources of measurement data to assess how a DDoS attack affects the several layers of the anycast deployment of Root DNS service. In addition, we are aware of no prior public studies on diverse anycast infrastructure operating under stress, including at the site and server level and its consequences on other services (collateral damage).

5. FUTURE WORK

Study of new events [49] can always provide new examples to strengthen our analysis. In addition, while we focus on IP anycast under stress, a full evaluation of Root DNS performance needs to consider the effects of caching and how recursive resolvers select and failover across different anycast services for the same DNS zone.

More important is to consider improving defenses. While additional anycast sites increase capacity, our work shows the importance of managing traffic across diverse sites (varying in capacity), since attackers are often unevenly distributed, and suggests potential directions for future improvements (§2.2).

6. CONCLUSIONS

This paper provides the first evaluation of anycast services under DDoS. Our work evaluates the Nov. 30 and Dec. 1, 2015 events on the Root DNS, evaluating the effects of those events on 10 different architectures, with most analysis based on publicly available data. Our analysis shows different behaviors across different letters (each a separate anycast services), at different sites of each letter, and at servers inside some sites. We identify the role of different policies at overloaded anycast sites: the choice to absorb attack traffic to protect other sites, or to withdraw service in hope that other sites can cover. We believe overall DNS service was robust to this attack, due to caching and the availability of multiple letters for service. However, we show that large attacks can overwhelm some sites of some letters. In addition, we show evidence that high traffic on one service can result in collateral damage to other services, possibly in the same data center. Our study shows the need to understand anycast design for critical infrastructure, paving the way for future study in alternative policies that may improve resilience.

Acknowledgments

The authors would like to thank Arjen Zonneveld, Jelte Jansen, Duane Wessels, Ray Bellis, Romeo Zwart, Colin Petrie, Matt Weinberg, Piet Barber, Alba Regalado, our shepherd Dave Levin, and the anonymous IMC reviewers for their valuable comments on paper drafts.

This research has been partially supported by measurements obtained from RIPE Atlas, an open measurements platform operated by RIPE NCC.

Giovane C. M. Moura, Moritz Müller and Cristian Hesselman developed this work as part of the SAND project (http://www.sand-project.nl).

Ricardo de O. Schmidt and Wouter de Vries' work is sponsored by the SAND and DAS (http://www.das-project.nl) projects.

John Heidemann and Lan Wei's work is partially sponsored by the U.S. Dept. of Homeland Security (DHS) Science and Technology Directorate, HSARPA, Cyber Security Division, via SPAWAR Systems Center Pacific under Contract No. N66001-13-C-3001, and via BAA 11-01-RIKA and Air Force Research Laboratory, Information Directorate under agreement numbers FA8750-12-2-0344 and FA8750-15-2-0224. The U.S. Government is authorized to make reprints

for Governmental purposes notwithstanding any copyright. The views contained in herein are those of the authors and do not necessarily represent those of DHS or the U.S. Government.

7. REFERENCES

[1] Abley, J., and Lindqvist, K. Operation of anycast services. RFC 4786, Internet Request For Comments, Dec. 2006. (also Internet BCP-126).

[2] Arbor Networks. Worldwide infrastructure security report, Sept. 2012. Volume VIII.

[3] Arbor Networks. Worldwide infrastructure security report, Jan. 2014. Volume IX.

[4] Arbor Networks. Rio Olympics Take the Gold for 540gb/sec Sustained DDoS Attacks! https://www.arbornetworks.com/blog/asert/rio-olympics-take-gold-540gbsec-sustained-ddos-attacks/, Aug. 2016.

[5] B-Root Operators. Personal communication, Dec. 2015.

[6] Ballani, H., and Francis, P. Towards a Global IP Anycast Service. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement* (Aug. 2007), pp. 301–312.

[7] Ballani, H., Francis, P., and Ratnasamy, S. A Measurement-based Deployment Proposal for IP Anycast. In *Proceedings of the ACM Internet Measurement Conference* (Oct. 2006), IMC, ACM, pp. 231–244.

[8] Bellis, R. Researching F-root Anycast Placement Using RIPE Atlas. https://labs.ripe.net/Members/ray_bellis/researching-f-root-anycast-placement-using-ripe-atlas, Oct. 2015.

[9] Beverly, R., Berger, A., Hyun, Y., and Claffy, K. Understanding the efficacy of deployed Internet source address validation filtering. In *Proceedings of the ACM Internet Measurement Conference* (Nov. 2009), IMC, ACM, pp. 356–369.

[10] Boothe, P., and Bush, R. Anycast Measurements Used to Highlight Routing Instabilities. NANOG 34, May 2005.

[11] Brownlee, N., Claffy, K., and Nemeth, E. DNS Root/gTLD Performance Measurement. In *Proceedings of the USENIX Large Installation System Administration conference* (Dec. 2001), pp. 241–255.

[12] Brownlee, N., and Ziedins, I. Response Time Distributions for Global Name Servers. In *Proceedings of the International conference on Passive and Active Measurements* (Mar. 2002), PAM.

[13] Bush, R. DNS Anycast Stability: Some Initial Results. CAIDA/WIDE Workshop, Mar. 2005.

[14] Bush, R., Karrenberg, D., Kosters, M., and Plzak, R. Root name server operational requirements. RFC 2870, Internet Request For Comments, June 2000. (also Internet BCP-40).

[15] Calder, M., Flavel, A., Katz-Bassett, E., Mahajan, R., and Padhye, J. Analyzing the Performance of an Anycast CDN. In *Proceedings of the ACM Internet Measurement Conference* (Oct. 2015), IMC, ACM, pp. 531–537.

[16] Castro, S., Wessels, D., Fomenkov, M., and Claffy, K. A Day at the Root of the Internet. *ACM Computer Communication Review 38*, 5 (Apr. 2008), pp. 41–46.

[17] Chirgwin, R. Linode: Back at last after ten days of hell. The Register, http://www.theregister.co.uk/2016/01/04/linode_back_at_last_after_ten_days_of_hell/, Jan. 2016.

[18] Chou, J. C.-Y., Lin, B., Sen, S., and Spatscheck, O. Proactive surge protection: a defense mechanism for bandwidth-based attacks. *IEEE/ACM Transactions on Networking (TON) 17*, 6 (Dec. 2009), pp. 1711–1723.

[19] Colitti, L. Effect of anycast on K-root. 1st DNS-OARC Workshop, July 2005.

[20] Czyz, J., Kallitsis, M., Gharaibeh, M., Papadopoulos, C., Bailey, M., and Karir, M. Taming the 800 Pound Gorilla: The Rise and Decline of NTP DDoS Attacks. In *Proceedings of the ACM Internet Measurement Conference* (Nov. 2014), IMC, ACM, pp. 435–448.

[21] Eastlake, D., and Andrews, M. Domain Name System (DNS) Cookies. RFC 7873 (Proposed Standard), May 2016.

[22] Elz, R., Bush, R., Bradner, S., and Patton, M. Selection and operation of secondary DNS servers. RFC 2182, Internet Request For Comments, July 1997. (also Internet BCP-16).

[23] Fan, X., Heidemann, J., and Govindan, R. Evaluating anycast in the Domain Name System. In *Proceedings of the IEEE Infocom* (Apr. 2013), IEEE, pp. 1681–1689.

[24] Ferguson, P., and Senie, D. Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing. RFC 2267, Internet Request For Comments, May 2000.

[25] Flavel, A., Mani, P., Maltz, D., Holt, N., Liu, J., Chen, Y., and Surmachev, O. Fastroute: A scalable load-aware anycast routing architecture for modern CDNs. In *12th USENIX Symposium on Networked Systems Design and Implementation* (May 2015), pp. 381–394.

[26] Fomenkov, M., Claffy, K., Huffaker, B., and Moore, D. Macroscopic Internet Topology and Performance Measurements From the DNS Root Name Servers. In *Proceedings of the USENIX Large Installation System Administration conference* (Dec. 2001), pp. 231–240.

[27] Gettys, J., and Nichols, K. Bufferbloat: dark buffers in the Internet. *Communications of the ACM 55*, 1 (Jan. 2012), pp. 57–65.

[28] Gillman, D., Lin, Y., Maggs, B., and Sitaraman, R. K. Protecting websites from attack with secure delivery networks. *IEEE Computer 48*, 4 (Apr. 2015), 26–34.

[29] H-Root Operators. Personal communication, Apr. 2016.

[30] Holterbach, T., Pelsser, C., Bush, R., and Vanbever, L. Quantifying interference between measurements on the RIPE Atlas platform. In *Proceedings of the ACM Internet Measurement Conference* (Oct. 2015), IMC, ACM, pp. 437–443.

[31] John, J. P., Moshchuk, A., Gribble, S. D., and Krishnamurthy, A. Studying spamming botnets using Botlab. In *Proceedings of the 6th USENIX Symposium on Network Systems Design and Implementation* (Boston, Massachusetts, USA, Apr. 2009), USENIX.

[32] Krebs, B. Israeli Online Attack Service 'vDOS' Earned $ 600,000 in Two Years http://krebsonsecurity.com/2016/09/israeli-online-attack-service-vdos-earned-600000-in-two-years/, Sept. 2016.

[33] Kührer, M., Hupperich, T., Rossow, C., and Holz, T. Exit from Hell? Reducing the Impact of Amplification DDoS Attacks. In *23rd USENIX Security Symposium* (Aug. 2014), pp. 111–125.

[34] Lee, B.-S., Tan, Y. S., Sekiya, Y., Narishige, A., and Date, S. Availability and Effectiveness of Root DNS servers: A long term study. In *Proceedings of the IEEE Network Operations and Management Symposium* (Apr. 2010), NOMS, pp. 862–865.

[35] Lee, T., Huffaker, B., Fomenkov, M., and Claffy, K. On the problem of optimzation of DNS root servers' placement. In *Proceedings of the International conference on Passive and Active Measurements* (Mar. 2003), PAM.

[36] Lentz, M., Levin, D., Castonguay, J., Spring, N., and Bhattacharjee, B. D-mystifying the D-root Address Change. In *Proceedings of the ACM Internet Measurement Conference* (2013), IMC, ACM, pp. 57–62.

[37] Liang, J., Jiang, J., Duan, H., Li, K., and Wu, J. Measuring Query Latency of Top Level DNS Servers. In *Proceedings of the International conference on Passive and Active Measurements* (Mar. 2013), PAM, pp. 145–154.

[38] Liu, Z., Huffaker, B., Fomenkov, M., Brownlee, N., and Claffy, K. Two Days in the Life of the DNS Anycast Root Servers. In *Proceedings of the International conference on Passive and Active Measurements* (Apr. 2007), PAM, pp. 125–134.

[39] Mockapetris, P. Domain names - implementation and specification. RFC 1035, Nov. 1987.

[40] Moura, G. C. M., de O. Schmidt, R., Heidemann, J., de Vries, W. B., Müller, M., Wei, L., and Hesselman, C. Nov. 30 datasets. http://traces.simpleweb.org/ and https://ant.isi.edu/datasets/anycast/, 2016.

[41] Perlroth, N. Tally of cyber extortion attacks on tech companies grows. New York Times Bits Blog, http://bits.blogs.nytimes.com/2014/06/19/tally-of-cyber-extortion-attacks-on-tech-companies-grows/, June 2016.

[42] ProtonMail. Guide to DDoS protection. https://protonmail.com/blog/ddos-protection-guide/, Dec. 2015.

[43] Ripe Atlas. Graphs: Probe firmware versions https://atlas.ripe.net/results/graphs/, Sept. 2016.

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

[45] 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.

[46] RIPE NCC Staff. RIPE Atlas: A global Internet measurement network. *The Internet Protocol Journal 18*, 3 (Sept. 2015), pp. 2–26.

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

[48] Root Server Operators. Events of 2015-11-30, Nov. 2015. http://root-servers.org/news/events-of-20151130.txt.

[49] Root Server Operators. Events of 2016-06-25, June 2016. http://root-servers.org/news/events-of-20160625.txt.

[50] Rossow, C. Amplification Hell: Revisiting Network Protocols for DDoS Abuse. In *Network and Distributed System Security (NDSS) Symposium* (Feb. 2014).

[51] RSSAC. Advisory on measurements of the Root Server System, Nov. 2014.

[52] Santanna, J. J., van Rijswijk-Deij, R., Hofstede, R., Sperotto, A., Wierbosch, M., Zambenedetti Granville, L., and Pras, A. Booters-An analysis of DDoS-as-a-service attacks. In *IFIP/IEEE Intl. Symposium on Integrated Network Management (IM)* (May 2015), IEEE, pp. 243–251.

[53] Sarat, S., Pappas, V., and Terzis, A. On the use of Anycast in DNS. In *Proceedings of the 15th International Conference on Computer Communications and Networks* (Oct. 2006), pp. 71–78.

[54] Shaikh, A., Kalampoukas, L., Dube, R., and Varma, A. Routing stability in congested networks: Experimentation and analysis. In *Proceedings of the ACM SIGCOMM Conference* (Aug. 2000), ACM, pp. 163–174.

[55] van Rijswijk-Deij, R., Sperotto, A., and Pras, A. DNSSEC and Its Potential for DDoS Attacks: a comprehensive measurement study. In *Proceedings of the ACM Internet Measurement Conference* (Nov. 2014), IMC, ACM, pp. 449–460.

[56] Vixie, P. Response Rate Limiting in the Domain Name System (DNS RRL). blog post http://www.redbarn.org/dns/ratelimits, June 2012.

[57] Weinberg, M., and Wessels, D. Review and analysis of anonmalous traffic to A-Root and J-Root (Nov/Dec 2015). In *24th DNS-OARC Workshop* (Apr. 2016). (presentation).

[58] Welzel, A., Rossow, C., and Bos, H. On Measuring the Impact of DDoS Botnets. In *7th European Workshop on System Security* (Apr. 2014).

[59] Wessels, D. Verisign's perspective on recent root server attacks. CircleID http://www.circleid.com/posts/20151215_verisign_perspective_on_recent_root_server_attacks/, Dec. 15 2015.

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

[61] Yan, H., Oliveira, R., Burnett, K., Matthews, D., Zhang, L., and Massey, D. BGPmon: A real-time, scalable, extensible monitoring system. In *Proceedings of the IEEE Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH)* (Mar. 2009), IEEE, pp. 212–223.

[62] Yu, Y., Wessels, D., Larson, M., and Zhang, L. Authority Server Selection in DNS Caching Resolvers. *SIGCOMM Computer Communication Review 42*, 2 (Mar. 2012), pp. 80–86.

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

[64] Zwart, R., and Buddhdev, A. Report: K-root on 30 November and 1 December 2015. RIPE Labs blog

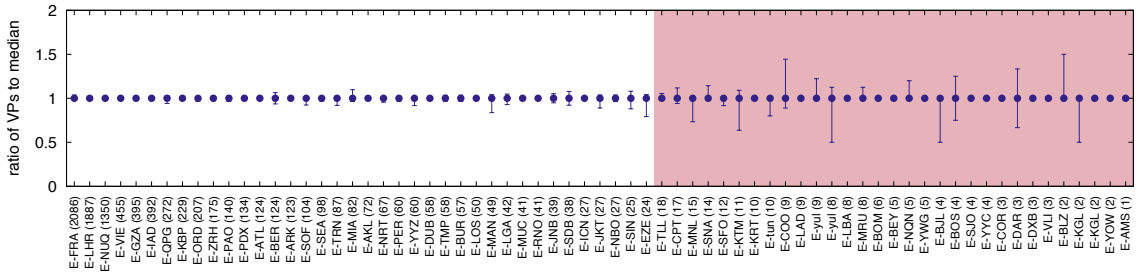
https://labs.ripe.net/Members/romeo_zwart/report-on-the-traffic-load-event-at-k-root-on-2015-11-30, Feb. 2015.

APPENDIX

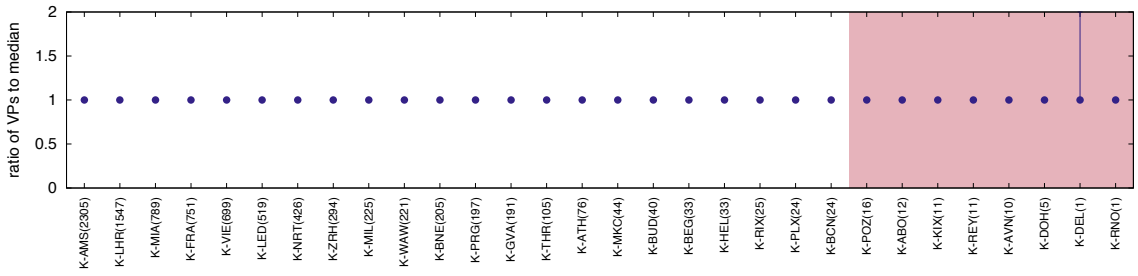
A. ADDITIONAL DATA FOR SITES

In §3.3 we showed that many sites for E- and K-Roots had greatly different catchments over the two days of the events (Figure 5). To demonstrate that those swings were correlated with events and not typical behavior, Figure 16 shows the same analysis for two normal days in the we after the events (2016-12-05 and 2016-12-06). We are aware of no abnormal traffic in these two days.

We see no large variance in catchments relative to their median for the sites in Figure 16 that see enough VPs to be statistically significant (median more than 20 VPs, the non-shaded region). K-Root (Figure 16b) shows no variation in catchment across all sites, and E-Root shows only minor variation in sites with a few VPs, indicating minor routing shifts. Combined with our analysis of individual VPs (§3.4), we think this make a strong case that the large swings in Figure 5 pertain directly to the events and are not noise.



(a) E-Root sites



(b) K-Root sites

Figure 16: Minimum and maximum number of VPs on the two normal days (2016-12-05 and 2016-12-06), normalized to median (shown between parenthesis per individual site), for sites from E- and K-Root. Counts are VPs that in each site's catchment over all 10-minute bins for two days, with sites ordered by median VPs.