A Day at the Root of the Internet

Sebastian Castro
CAIDA and NIC Chile
secastro@caida.org

Duane Wessels
CAIDA and The Measurement Factory, Inc.
wessels@measurement-factory.com

Marina Fomenkov, Kimberly Claffy
CAIDA, University of California San Diego
marina@caida.org, kc@caida.org

ABSTRACT

We analyzed the largest simultaneous collection of full-payload packet traces from a core component of the global Internet infrastructure ever made available to academic researchers. Our dataset consists of three large samples of global DNS traffic collected during three annual "Day in the Life of the Internet" (DITL) experiments in January 2006, January 2007, and March 2008. Building on our previous comparison of DITL 2006 and DITL 2007 DNS datasets [28], we venture to extract historical trends, comparisons with other data sources, and interpretations, including traffic growth, usage patterns, impact of anycast distribution, and persistent problems in the root nameserver system that reflect ominously on the global Internet. Most notably, the data consistently reveals an extraordinary amount of DNS pollution – an estimated 98% of the traffic at the root servers should not be there at all. Unfortunately, there is no clear path to reducing the pollution, so root server operators, and those who finance them, must perpetually overprovision to handle this pollution. Our study presents the most complete characterization to date of traffic reaching the roots, and while the study does not adequately fulfill the "Day in the Life of the Internet" vision, it does succeed at unequivocally demonstrating that the infrastructure on which we are all now betting our professional, personal, and political lives deserves a closer and more scientific look.

> *About 30 years ago there was much talk that geologists ought only to observe and not theorise; and I well remember saying that at this rate a man might as well go into a gravel pit and count the pebbles and describe the colors. How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service.*
> *– Charles Darwin to W.W. Bates, 22 Nov 1860*

Categories and Subject Descriptors

C.2.2 [**Network protocols**]: Applications—DNS; C.2.3 [**Network operations**]: Network management; C.2.4 [**Distributed Systems**]: Client/server; C.2.5 [**Local and Wide-Area Networks**]: Internet

General Terms

Measurement, Management, Human Factors, Legal Aspects

Keywords

Domain Name System, misconfiguration, Day in the Life of the Internet, root servers, traffic growth, unwanted traffic

1. INTRODUCTION

When we first read the recommendation to "capture a day in the life of the Internet" in a National Academies workshop report [3, 9], we ignored it as a glib suggestion from researchers who made their living studying computer systems presumably contained within a single room, and who assumed Internet measurement was mostly a technical problem, rather than primarily deterred by issues of economics, ownership, and trust. But an even more far-fetched recommendation pursued in 2005 [2] brought the DITL idea back into focus.[1] Since we believed GENI would fail for the same reasons a "Day in the Life" experiment would fail, i.e., policy rather than technology challenges, we decided that failing at the latter would be less expensive to U.S. taxpayers, but would succeed at raising awareness of the policy problems still underappreciated by the research community.

By 2005, CAIDA had been funded by NSF to study the Domain Name System (DNS) [23] for several years. As part of this research CAIDA responded to the Root Server System Advisory Committee's invitation to help DNS root operators study and improve the integrity of the root server system. On the few years of trust we had built with these operators, we asked them to participate in a simultaneous collection of a day of traffic to (and in some cases from) the DNS root name servers. Less ambitious than capturing data from the entire Internet, but it was a strategic place to start for both technical and policy reasons. It made sense technically because the DNS [23] is a fundamental component of today's Internet, mapping domain names used by people and their corresponding IP addresses. The data for this mapping is stored in a tree-structured distributed database where each nameserver is authoritative for a part of the naming tree, and the *root nameservers* play a vital role providing authoritative referrals to nameservers for all top-level domains, which recursively determine referrals for all host names on the Internet, among other infrastructure information. Measuring the roots also made sense from a policy perspective, because the DNS root servers are not faced with the same counterincentives or legal barriers to sharing data as commercially provisioned Internet infrastructure, so there was an opportunity to break the impasse typically associated with sharing interesting data about the Internet with researchers.

Three anycasted root servers (C, F, and K) participated in DITL in January 2006,[2] four anycasted root servers (C, F, K, M) in Jan-

[1] GENI was a proposal by academic researchers to build a testbed for a future Internet, because they were not allowed to "innovate" in the operational infrastructure that the current Internet had become.

[2] Picking the second week in January was a mistake in retrospect, but it was not a funded project at the time so we used the first spare week we had. A bigger mistake was trying to repeat these dates in 2007 to be consistent. This year we moved it to March.

uary 2007, and eight root servers (plus several authoritative TLD servers) in March 2008. We first give some background on the root servers, describe the data, then extend the results of our previous comparison of 2006 and 2007 collected data, including deeper analysis of the pollution, and finally offer some interpretations of the data and the experiment itself.

2. BACKGROUND ON ROOT SERVERS

The original DNS design provisioned only 13 root nameservers ("roots") to provide the bootstrap foundation for the entire database. Explosive Internet growth challenged this limitation, while also increasing the cost of any transition to a new system. With more concern than consensus regarding its potential impact, anycast [13] was finally deployed in 2002 to allow more than the static 13 root-servers to serve the root zone without changing the existing protocol. As of mid-2008, 7 of the 13 root nameservers use anycast deployment: C, F, I, J, K, L and M [25]. (For an explanation of anycast and the meaning of *global* and *local* instances, please refer to Liu *et al.* [21].)

Deployment of anycast technology made the root servers more resilient to DDOS attacks [15] by limiting the spread of attack traffic beyond the root server nearest to the attack source(s). Anycast also greatly expanded the deployment of root servers around the world, improving service in areas previously underserved [19]. Performance improvements experienced by particular clients depend on BGP path selection, which itself depends on economic and policy considerations as well as network dynamics. Anycast deployment also rendered the measurement problem even more challenging, as machines composing a single root server may be anycasted across dozens of fundamentally different legal jurisdictions. Previous DITL experiments enabled us to study the impact of anycast on the stability of DNS root name service, both within and across anycast clouds [21, 28], during which we found that the anycast deployment appears to be stable, efficient, and enabling better service to the worldwide population of users.

3. DATA

	2007 Roots	2008 Roots
Dataset duration	24 h	24 h
Dataset begin	January 9, noon (UTC)	March 19 midnigth (UTC)
# of instances: observed/total X_L: local anycast X_G: global anycast X_U: unicast	C_G: 4/4 F_G: 2/2 F_L: 34/38 K_G: 5/5 K_L: 10/12 M_G: 6/6	A_U: 1/1 C_G: 4/4 E_U: 1/1 F_G: 2/2 F_L: 38/40 H_U: 1/1 K_G: 5/5 K_L: 10/12 M_G: 6/6
Query Count	3.84 Billions	8.0 Billions
Unique clients	2.8 Millions	5.6 Millions
Recursive Queries	17.04%	11.99%

Table 1: General statistics of the 2007 and 2008 datasets

CAIDA and the DNS-OARC [11] have now completed three annual Day-in-the-Life DNS measurement experiments, with increasing participation each year. On January 10–11, 2006, we coordinated concurrent measurements of three DNS root server anycast clouds (C, F, and K, see [21] for results and analysis). On January 9–10, 2007, four root servers (C, F, K, and M) participated in simultaneous capture of packet traces from almost all instances of their anycast clouds [8]. On March 18–19, 2008, operators of eight root servers (A, C, E, F, H, K, L and M), five TLDs (.ORG, .UK, .BR, .SE and .CL), two RIRs (APNIC and LACNIC), and seven operators of project AS112 joined this collaborative effort. Two ORSN servers, B in Vienna and M in Frankfurt, participated in our 2007 and 2008 collection experiments. **To the best of our knowledge, these events deliver the largest scale simultaneous collection of full-payload packet traces from a core component of the global Internet infrastructure ever made available to academic researchers**. DNS-OARC provides limited storage and compute power for researchers to analyze the DITL data, which for privacy reasons cannot leave OARC machines.[3]

Building on our previous comparison of 2006 and 2007 data [28], we focus on changes between 2007 and 2008. Table 1 shows general statistics of the traces. From our 48 hours of data in each year, we selected for analysis the 24-hour interval with the most complete coverage: starting at noon on January 9, 2007 [16], and starting at midnight March 19, 2008 [17].

4. WHAT DID WE LEARN?

In this section we highlight interesting aspects of the DITL data that both confirm and extend results of previous work [10, 34, 32, 21, 8, 28].

4.1 Query Workload

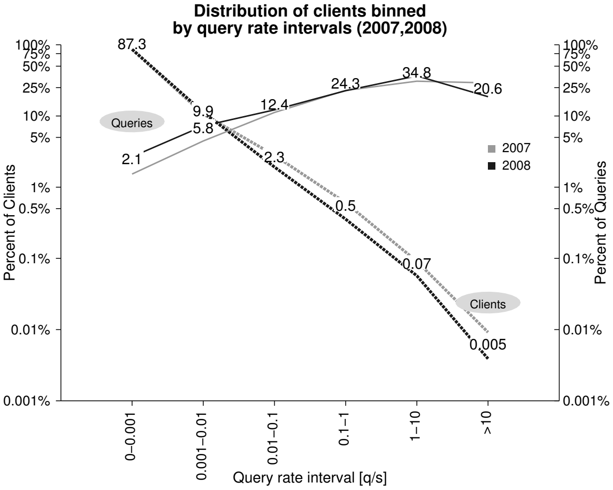

Figure 1: Distribution of clients by mean query rate in 2007 and 2008. The Y-axes use a log scale. This graph shows just a few clients (two rightmost categories) are responsible for more than 50% of the traffic at the observed root servers.

Clients querying root servers exhibit surprising variation in query rates, from one query in 24 hours (1.09M or 20% of observed clients in 2008) to tens of millions of queries in the same period (19 clients in 2008). In our 2007 sample, 438K (17%) clients sent just one query, while 10 exceeded the ten million barrier.

Figure 1 shows the binned distribution of average query rate for clients. Each bin is one order of magnitude wide, and the overall range is from <0.001 q/s to >10 q/s. The decreasing lines show the distribution of client query rates; the increasing lines show the distribution of total query load produced by clients as a function

[3] OARC also hosts equipment for researchers who need more compute or storage than OARC can provide.

of their query rate. Both distributions are highly skewed. The two groups with the lowest query rates (less than 1 query per 100s) contain about 97% of clients, while contributing slightly less than 6% of the observed query load (10.3% during 2008). On the opposite end we see that <0.1% of the busiest clients, which compose the two groups with the highest query rates (greater than 1 query per second), are responsible for 60% of the total load in the 2007 dataset and 55% of the total load in the 2008 dataset. As we will see in Section 4.3, these busier clients are more likely to be generating mostly invalid queries.

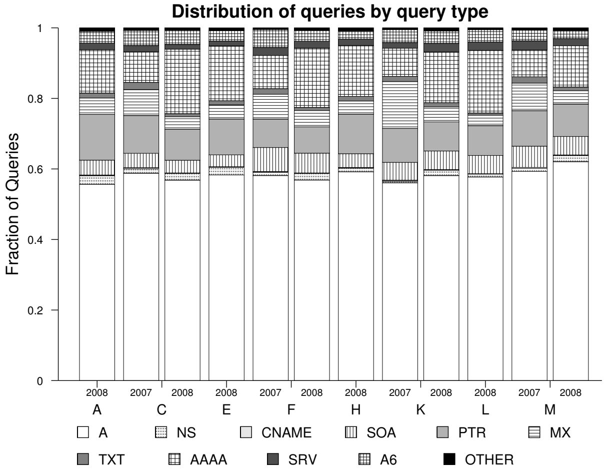

Figure 2: Distribution of queries by type. The distribution of query type shifted in 2008, including a significant increase in AAAA queries due to the addition of IPv6 glue records to six root servers in February 2008.

Figure 2 shows the distribution of queries by type received at each root. For all root servers measured, A-type queries, used to request an IPv4 address for a given hostname, are the most common (about 60% of the total), and their fraction has remained stable for the last three years. Of note in Figure 2 is the even greater increase than last year of AAAA-type queries, which map an IPv6 address to a hostname, on all four root servers. Last year we attributed the slight increase to more clients using operating systems with native IPv6 capabilities such as Apple's MacOSX and Microsoft's Windows Vista [28]. This year the increase is more substantial, from around 8% on average to 15%. At least some of this increase is attributable to the addition of IPv6 glue records to six of the root servers during February 2008 [1].

4.2 Traffic growth

Figure 3 shows the average query rate received by each root. For four roots we have three years of data, which allows inference of trends. For these roots, query rates doubled on average from 2006 to 2007: 70% growth at C-root, 51% at F-root, and 31% at K-root. There jump between January 2007 and March 2008 was smaller (40%, 13%, and 33% and 5% for C, F, K and M respectively), consistent with other reports that Internet traffic growth is itself slowing [24]. The crossover split is 95/5, that is, 5% of the clients contribute 95% of the query workload, a common crossover value for IP traffic [7]. Traffic growth by individual instance is highly variant, and 19 of the 54 instances common in 2007 and 2008 actually decreased their average query rate in the 2008 data. Of the other 35 instances, 17 increased average query load less than 50%, 18 increased more than 50%, and 9 increased over 100% (Traffic to f-root's Caracas node multiplied 19-fold, since it was just getting started.).

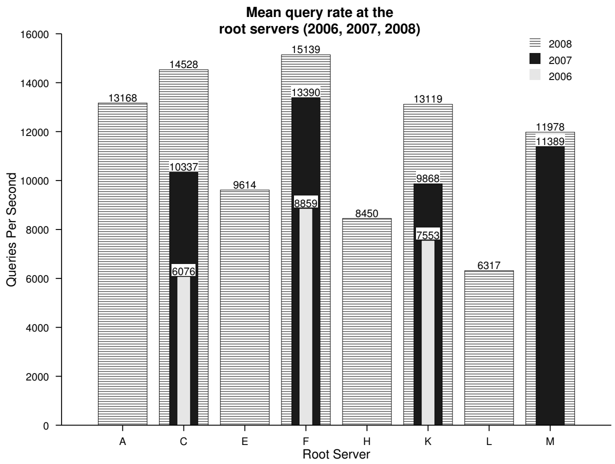

Figure 3: Average query rate observed at root servers participating in DITL 2006, 2007 and 2008. The solid light, dark, and hashed bars represent the average query rates in 2006, 2007, 2008, respectively.

4.3 Validity of Traffic

For years, an extraordinary percent (over 90%) of queries reaching the roots have been invalid, and DITL data suggests that since 2007 the pollution has grown faster than the legitimate traffic. We categorize queries according to the nine types of invalid queries established in [34]:
1. *unused query class*, a query not of the standard five classes [5];
2. *A-for-A*, a query of type A where the query "name" is already an IPv4 or IPv6 address;
3. *invalid TLD*, a query for a name with an invalid TLD [14];
4. *non-printable characters*, a query for a name with characters that are not alphanumeric or dash;
5. *queries with _*, to show the wide use of the invalid _ symbol [23];
6. *RFC 1918 PTR*, a PTR query for an IP address from private address space;
7. *Identical queries*, having the same type, class, name and ID;
8. *Repeated queries*, having the same type, class, and name, but different ID's;
9. *Referral-not-cached* queries repeated because the previous referral was not cached.
Queries that do not fall into any of those nine categories are considered valid.[4]

Table 2 categorizes a 10% sample of unique clients querying the root servers in the 2008 data.[5] Keeping the state necessary to identify Identical, Repeated and Referral-not-cached queries for the whole data set is computationally prohibitive on the computing infrastructure we have at OARC, where the data must remain.

In pursuit of deeper insight into the nature of this pollution, Figure 4 shows the distribution of valid and invalid queries vs. query rates for the A, C, E, F, H, K, L, and M root servers. The percentage of valid queries (the black at the top of each bar) is distinctly anticorrelated with query rate: the higher the query rate of a client, the lower its fraction of legitimate queries. The rightmost three groups,

[4] We also investigated the presence of repeated queries due to "glue record refresh", a behavior of BIND 9 and DJBDNS that could be misclassified as repeats, but we found that such behavior accounted for at most 1% of the queries in our subsample.

[5] The 2007 results (not shown) are similar, although slightly higher fraction of legitimate queries than in 2008.

Category	A	C	E	F	H	K	L	M	Total
Unused query class	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.1	0.1
A-for-A	1.6	1.9	1.2	3.6	2.7	3.8	2.6	2.7	2.7
invalid TLD	19.3	18.5	19.8	25.5	25.6	22.9	24.8	22.9	22.0
non-printable char.	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.0
queries with _	0.2	0.1	0.2	0.1	0.2	0.1	0.1	0.1	0.1
RFC 1918 PTR	0.6	0.3	0.5	0.2	0.5	0.2	0.1	0.3	0.4
identical queries	27.3	10.4	14.9	12.3	17.4	17.9	12.0	17.0	15.6
repeated queries	38.5	51.4	49.3	45.3	38.7	42.0	44.2	43.9	44.9
referral-not-cached	10.7	15.2	12.1	10.9	12.9	11.1	14.3	11.1	12.4
Valid	1.7	2.0	1.8	1.9	1.8	2.0	1.8	1.8	1.8
Valid 2006		2.3		2.1		2.5			2.1
Valid 2007		4.1		2.3		1.8		4.4	2.5

TLD	2007	2008
local	5.018	5.098
belkin	0.436	0.781
localhost	2.205	0.710
lan	0.509	0.679
home	0.321	0.651
invalid	0.602	0.623
domain	0.778	0.550
localdomain	0.318	0.332
wpad	0.183	0.232
corp	0.150	0.231

Table 2: a) Taxonomy of queries (in %) based on 2008 data; b) Percentage of queries (in %) for top ten invalid TLD using 2007 and 2008 data.

corresponding to rates of more than 1 query per 10 second, contain practically no legitimate queries! Unfortunately, these groups contribute more than 78% of the load on each anycast cloud. Therefore, the overall percentage of valid queries at root servers remains alarmingly low, and dropping, now below 2.0% of the total, from 2.5% observed in the 2007 data (cf. Table 2).

4.4 Sources of spurious traffic

Misconfigurations and mistakes can be as costly as targeted DoS attacks, especially because attacks on root servers have thus far been temporary events [15] while misconfigured clients continuously bombard the roots at the rate of thousands per second, and the roots must continually upgrade their infrastructure to accommodate this growth, the vast majority of which is pollution. We have previously published papers [32] and tools [31] to educate and assist DNS operators with repairing broken configurations that emit polluting queries, but as with other sources of DNS pollution [6], there is insufficient incentive to invest in repair, or even understanding the root of the problem. We next give details on some the most common categories of invalid queries seen at the roots.

4.4.1 A-for-A queries

In the DITL 2008 root server traces, we found around 334K clients (6% of the total) sent at least one A-for-A query, 3031 clients had more than 80% of their queries in this category, and one client sent 3.9M A-for-A queries! RFC 4697 [18] notes this type of query can result from poorly configured stub resolvers. Four patterns of this type of error stand out: address sweeping (22% of queries);[6] A-for-A6/A-for-AAAA queries (3.3%); queries for RFC 1918 address space (2.3%); and queries for addresses on the same /24 of the source address (0.31%).

4.4.2 Recursive queries

Since the root servers do not provide recursion, any recursive query sent to them will get a referral answer rather than a direct response. We observed in 2008 that 1.97M clients (36.4% of the total) sent at least one recursive query to the roots (290K did so in 2007, around 11.3% of the total). Recursive queries are not necessarily errors, since command line programs and DNS APIs (such as Net::DNS for Perl) may be set by default to send recursive queries, conceivably directed at the root for diagnosis purposes. Of the 1.97M clients sending recursive queries, 1.1M sent at most 5 queries, all of them recursive, suggesting diagnostic traffic. If we exclude these diagnostic queries from the pollution category, it negligibly reduces the level of pollution by 0.2%.

4.4.3 Invalid TLD

In 2007 the number of queries reaching the roots for invalid TLDs reached 20.6% and in 2008 increased to 22%. Table 2b shows the top 10 invalid TLDs observed: `.local` is at the top both years, with around 5% of the total queries. These queries are produced by stub resolvers with misconfigured search paths, and caching resolvers leaking these queries to the Internet. RFC 2606 declared 4 TLD's *reserved*, of which `.localhost` and `.invalid` are included in the top-10 list of Table 2b. To mitigate this category of pollution, IANA could add the commonly queried invalid TLDs such as `.local`, `.lan`, `.home`, `.domain`, `.localdomain` to the list of reserved top level domains, and the DNS community could standardize cache implementations to keep such queries local (either with valid or error responses). Although such changes would substantially reduce unwanted traffic at the roots, the Internet standards community's performance at keeping local traffic local has not been stellar [6].

4.4.4 No correlation with DNS blacklists

In the past decade operators have increasingly used realtime DNS-based blacklists to filter spam and other types of unwanted traffic. We wondered if these blacklists correlated with pollution at the roots, i.e., do prefixes/ASes containing IPs listed in these blacklists contribute a variety of unwanted traffic including unnecessary DNS pollution to the root servers? Our results were inconclusive, but in general blacklisted prefixes seem uncorrelated with those sending lots of queries to the roots.

4.5 Source Port Randomization Quality

In early 2008, Dan Kaminsky discovered a DNS protocol vulnerability that allows an attacker to poison a cache in less than ten seconds [29]. The exploit takes advantage of the typically poor or no Source Port Randomization in most DNS cache resolver implementations. The exploit was not made public until July, three months after the DITL 2008 collection, but we used the traces collected at three TLDs participating in DITL (.ORG, .UK and .BR) to estimate the quality of source port randomization before the announcement. We used two randomness metrics suggested by CERT.at [20] and one implemented in a public service tool by Wessels [12, 33] on the clients in the 2008 data. The metrics gave similar results, estimating that 50–64% of clients were using poor randomization, with the rest roughly divided between good and mediocre.

[6] Around 18K clients on different networks sent apparently coordinated A-for-A queries for each address in four sequential /8 prefixes from 80/8 to 83/8, and a subset of around 8K clients also sent the same type of queries for other two /8 prefixes from 88/8 to 89/8, fixing the last octet and iterating over the third octet as if trying to evade detection.

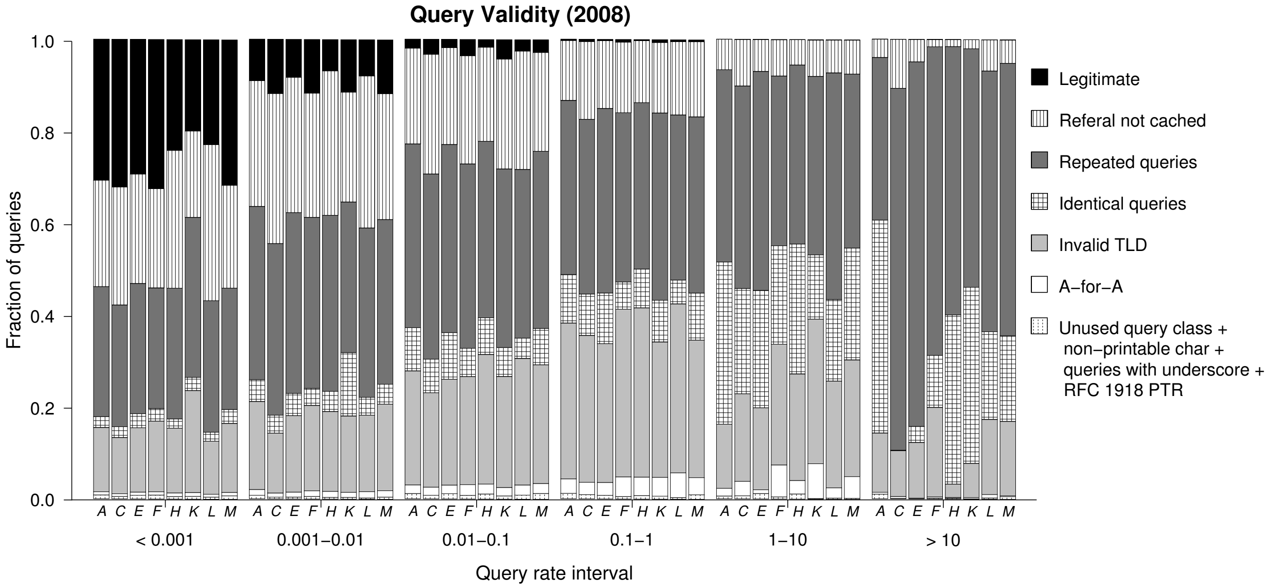

Figure 4: Validity of queries (2008). A strong correlation between the query rate and the amount of unwanted traffic could be observed. The amount of legitimate traffic (in black) reduces when the query rate increases.

4.6 Related work

The DNS has been studied since its inception [23], from a variety of angles from workload to topology to performance to security. Studies on improving the integrity of the DNS root system are less common. Malone [22] proposed that busy cache resolvers become slaves of the root zone to reduce the response time, which would be particularly efficient in cases when the queried name does not exist. This solution would also eliminate some of the pollution at the root servers.

Pappas *et al.* [27] studied how three operational errors (lame delegation, diminished server redundancy, and cyclic zone dependency) can affect the robustness of DNS, and operational choices in one zone can affect the availability of other zones. They suggest that DNS include a systematic checking mechanism to cope with these operational errors. In more recent work [26] these authors proposed a distributed troubleshooting tool to help administrators find DNS configuration errors that are not visible from a single vantage point. van Wanrooij *et al.* [30] took a different approach, evaluating the configuration of the .NL zone based on the policy used by the administrator of the .NL zone (SIDN, Netherlands) [27]. They showed that 14% of the zones presented did not meet all SIDN configuration requirements. Given that the .NL zone is considered relatively well-managed, we expect this number is higher for other TLDs.

5. INTERPRETATIONS

Although we were skeptical regarding having any success with a global “Day in the Life of the Internet” attempt, we did better than we cynically expected. By relying on our trust relationship developed with the root name servers, we were able to capture and analyze the richest data set about a core component of the global Internet infrastructure – the DNS root name servers – ever made available to academic researchers, and through NSF’s investment in OARC ensure that other researchers would also have access to this data (at no cost, if they fill out the OARC paperwork). The DITL data correlates with other sources of traffic growth data [24], including the recent slowdown in traffic growth, although there is much more to explore. We have also detected non-trival (nor strong) signs of IPv6 growth reflected at the root nameservers.

We are trying to help the root operators understand more about the causes of and opportunities to mitigate the vast and increasing fraction of inappropriate traffic at the roots. We know that repeated, identical and referral-not-cached queries constitute 73% of the total load, and we know this imposes an unnecessary burden on both technological and human resources of root server operators. We also now know that among clients, the higher the query rate, the lower the fraction of valid queries. We believe those queries are produced by a combination of poorly designed cache implementations, bad network practice implementations (packet filters not allowing the response to reach the initiator) and zone configuration errors. But without a more precise understanding of what is causing this pollution, we have little hope of mitigating it, or the perpetual provisioning expense it incurs on DNS root operators. Other than Wessels *et al.* [35], no work has developed a model of a well-behaving cache resolver or typical DNS traffic sources. The research community still needs a model of baseline DNS behavior to improve our ability to detect misbehavior in the future. (Just like we need baseline models for for non-DNS traffic.)

We invite others to analyze this data, and not just because it has the property of being available to researchers, although such a rare property should not be discounted. The DITL DNS data reveals some of the more dire statistics about the global Internet [4], and is likely to point if not lead researchers toward other telling indicators. We may not know what a day in the life of the Internet looks like yet, but we have unequivocally demonstrated that the infrastructure on which we are all now betting our professional, personal, and political lives deserves a closer scientific look. What remains to be seen is how we get there.

Acknowledgments. We thank ISC, RIPE, WIDE, Cogent, Verisign, NASA, United States Army, ICANN, Registro.br, Afilias, Nominet, NIC Chile, IIS.SE, Home of the Brave GmbH, and Lynix for collecting data. P. Vixie, K. Mitchell, Joao Damas and L. Bicknell from ISC helped with data storage and explained F-root’s anycast deployment. This work was supported by NSF grant OCI-0427144.

6. REFERENCES

[1] IPv6 Address Added for Root Servers in the Root Zone. http://www.icann.org/en/announcements/announcement-04feb08.htm.

[2] Global Environment for Network Investigation (now ”innovations”), 2005. http://www.geni.net.

[3] A Day in the Life, 2006. http://blog.caida.org/best_available_data/2006/09/04/a-day-in-the-life/.

[4] Top ten things lawyers should know about Internet research, #3, 2008. http://blog.caida.org/best_available_data/2008/04/18/top-ten-things-lawyers-should-know-about-internet-research-3/.

[5] D. E. 3rd, E. Brunner-Williams, and B. Manning. Domain Name System IANA Considerations, 2000. http://www.rfc-editor.org/rfc/rfc2929.txt.

[6] A. Broido, Y. Hyun, M. Fomenkov, and K. Claffy. The windows of private DNS updates. *SIGCOMM Comput. Commun. Rev.*, 36(3):93–98, 2006.

[7] A. Broido, Y. Hyun, R. Gao, and K. Claffy. Their share: diversity and disparity in IP traffic. In *PAM 2004 Proceedings*, pages 113–125, 2004.

[8] CAIDA and DNS-OARC. A Report on DITL data gathering Jan 9-10th 2007. http://www.caida.org/projects/ditl/summary-2007-01/.

[9] N. R. Council. *Looking over the Fence: A Neighbor’s View of Networking Research.* National Academies Press, 2001.

[10] P. B. Danzig, K. Obraczka, and A. Kumar. An analysis of wide-area name server traffic: a study of the Internet Domain Name System. *SIGCOMM Comput. Commun. Rev.*, 22(4):281–292, 1992.

[11] DNS-OARC. Domain Name System Operations, Analysis, and Research Center. https://www.dns-oarc.net/.

[12] Duane Wessels. https://www.dns-oarc.net/oarc/services/dnsentropy.

[13] T. Hardie. Distributing Authoritative Nameservers via Shared Unicast Addresses, 2002. http://www.ietf.org/rfc/rfc3258.txt.

[14] IANA. List of valid TLD. http://data.iana.org/TLD/tlds-alpha-by-domain.txt.

[15] ICANN. Root servers attack factsheet, 2007. http://www.icann.org/announcements/factsheet-dns-attack-08mar07.pdf.

[16] C. D. in the Life of the Internet (DITL) Project Team. Day in the Life of the Internet, January 9-10, 2007 (DITL-2007-01-09) (collection). http://imdc.datcat.org/collection/1-031B-Q=Day+in+the+Life+of+the+Internet%2C+January+9-10%2C+2007+%28DITL-2007-01-09%29.

[17] C. D. in the Life of the Internet (DITL) Project Team. Day in the Life of the Internet, March 18-19, 2008 (DITL-2008-03-18) (collection). http://imdc.datcat.org/collection/1-05MM-F=Day+in+the+Life+of+the+Internet%2C+March+18-19%2C+2008+%28DITL-2008-03-18%29.

[18] M. Larson and P. Barber. Observed DNS Resolution Misbehavior. http://www.ietf.org/rfc/rfc4697.txt.

[19] T. Lee, B. Huffaker, M. Fomenkov, and K. Claffy. On the problem of optimization of DNS root servers’ placement. In *PAM 2003 Proceedings*, 2003.

[20] O. Lendl and L. A. Kaplan. Patching Nameservers: Austria reacts to VU#800113. http://cert.at/static/cert.at-0802-DNS-patchanalysis.pdf.

[21] Z. Liu, B. Huffaker, N. Brownlee, and K. Claffy. Two Days in the Life of the DNS Anycast Root Servers. In *PAM 2007 Proceedings*, pages 125–134, 2007.

[22] D. Malone. The root of the matter: hints or slaves. In *IMC ’04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 15–20, New York, NY, USA, 2004. ACM.

[23] P. Mockapetris. Domain Names - Concepts and Facilities, 1987.

[24] A. Odlyzko. Minnesota Internet Traffic Studies (MINTS). http://www.dtc.umn.edu/mints/home.php.

[25] R. S. Operators. Root Server Technical Operations. http://www.root-servers.org/.

[26] V. Pappas, P. Fältström, D. Massey, and L. Zhang. Distributed DNS troubleshooting. In *NetT ’04: Proceedings of the ACM SIGCOMM workshop on Network troubleshooting*, pages 265–270, New York, NY, USA, 2004. ACM.

[27] V. Pappas, Z. Xu, S. Lu, D. Massey, A. Terzis, and L. Zhang. Impact of configuration errors on DNS robustness. *SIGCOMM Comput. Commun. Rev.*, 34(4):319–330, 2004.

[28] S. Castro, D. Wessels, and Kimberly Claffy. A Comparison of Traffic from the DNS Root Nameservers as Measured in DITL 2006 and 2007. http://www.caida.org/research/dns/roottraffic/comparison06_07.xml.

[29] US-CERT. Vulnerability note VU#800113: Multiple DNS implementations vulnerable to cache poisoning. http://www.kb.cert.org/vuls/id/800113.

[30] W. van Wanrooij and A. Pras. DNS Zones Revisited. *Open European Summer School and IFIP WG6.4/6.6/6.9 Workshop (EUNICE)*, 2005.

[31] D. Wessels. *dnstop*. http://www.caida.org/tools/utilities/dnstop/.

[32] D. Wessels. Is your caching resolver polluting the internet? In *NetT ’04: Proceedings of the ACM SIGCOMM workshop on Network troubleshooting*, pages 271–276, New York, NY, USA, 2004. ACM.

[33] D. Wessels. Measuring DNS Source Port Randomness. First CAIDA/WIDE/CASFI workshop, August 2008. http://www.caida.org/workshops/wide/0808/slides/source_port_randomness.pdf.

[34] D. Wessels and M. Fomenkov. Wow, That’s a lot of packets. In *PAM 2002 Proceedings*, 2002.

[35] D. Wessels, M. Fomenkov, N. Brownlee, and K. Claffy. Measurements and Laboratory Simulations of the Upper DNS Hierarchy. In *PAM 2004 Proceedings*, pages 147–157, 2004.