Clouding up the Internet: how centralized is DNS traffic becoming?

Giovane C. M. Moura (1) Sebastian Castro (2) Wes Hardaker (3)
Maarten Wullink (1) Cristian Hesselman (1,4)
1: SIDN Labs 2: InternetNZ 3: USC/ISI 4: University of Twente

ABSTRACT

Concern has been mounting about Internet centralization over the few last years – consolidation of traffic/users/infrastructure into the hands of a few market players. We measure DNS and computing centralization by analyzing DNS traffic collected at a DNS root server and two country-code top-level domains (ccTLDs) – one in Europe and the other in Oceania – and show evidence of concentration. More than 30% of all queries to both ccTLDs are sent from 5 large cloud providers. We compare the clouds' resolver infrastructure and highlight a discrepancy in behavior: some cloud providers heavily employ IPv6, DNSSEC, and DNS over TCP, while others simply use unsecured DNS over UDP over IPv4. We show one positive side to centralization: once a cloud provider deploys a security feature – such as QNAME minimization – it quickly benefits a large number of users.

ACM Reference Format:
Giovane C. M. Moura, Sebastian Castro, Wes Hardaker, Maarten Wullink, Cristian Hesselman. 2020. Clouding up the Internet: how centralized is DNS traffic becoming?. In *ACM Internet Measurement Conference (IMC '20), October 27–29, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3419394.3423625

1 INTRODUCTION

There have been growing concerns over the last few years about a phenomenon described as Internet centralization and consolidation [6–8, 23, 41], which is the trend towards increasing the concentration of traffic/infrastructure/users/services into the control of a small set of companies.

Centralization poses various risks [39, 41]. From a technical point-of-view, such concentration in the hands of few market players may lead to large single points-of-failure [2, 6]. This was demonstrated during the two large scale Denial-of-Service (DoS) attacks against two large authoritative DNS service providers (Dyn in 2016 [33] and AWS in 2019 [46]). In both cases, parts of their DNS infrastructure became unreachable and, consequently, their clients lost connectivity with each service. In the case of the DDoS against Dyn, many prominent websites became at least partially unreachable, including Netflix, Twitter, and The New York Times.

IMC '20, October 27–29, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8138-3/20/10…$15.00
https://doi.org/10.1145/3419394.3423625

Company	ASes	Public DNS?
Google	15169	Yes
Amazon	7224, 8987, 9059, 14168, 16509	No
Microsoft	3598,6584, 8068–8075, 12076, 23468	No
Facebook	32934	No
Cloudflare	13335	Yes

Table 1: Cloud/content providers and their ASes

Among the large service providers, we see a handful of companies dominating the cloud computing industry, offering on-demand services such as storage and data processing. Given their market dominance, in this paper we investigate how the *market dominance* of five large cloud/content providers (Table 1) translates into *DNS traffic* dominance. We analyze the three largest cloud providers (Amazon, Google, and Microsoft) and Cloudflare (given they run a large public DNS service [1]). For comparison, we also add the hyper-giant Facebook, for being the largest social networking platform – which also employs a CDN to operate efficiently. While there are other cloud providers, we focus on this subset given they are either among the largest and/or also run public DNS services.

To study the growth of cloud providers, we select the DNS protocol [25] for its core (ironically centralized) role in connecting applications, services, and hosts. Studying DNS queries acts as a proxy metric for measuring the usage and popularity of requested resources on the Internet.

Using DNS vantage points, we analyze traffic collected at the authoritative DNS servers [20] of the one of the Root DNS root servers [37] (B-Root) and two country-code top-level domains (ccTLDs): The Netherlands' `.nl` (located in Europe with ~6 million domain names) and New Zealand's `.nz` (located in Oceania, with more than 700 thousand domain names). Comparing traffic from the entire Internet to a root server and those from two different countries with different official languages allows us to determine if the same trends can be seen from multiple vantage points.

We make the following contributions: First, we measure DNS centralization within the Internet today, from the point-of-view of ccTLDs (§4) and a DNS root server. Secondly, we analyze and compare the query patterns from these cloud providers (CPs hereafter), and their resolver infrastructure. We show that despite being large, we see that they are far from being homogeneous with regard adoption of latest protocols, and show large diversity in terms IPv6, DNSSEC, and QNAME minimization deployment.

2 DNS SERVERS AND CCTLDS

Two types of infrastructure make up DNS service on the Internet. *DNS authoritative servers* serve the content of the DNS zones [20] on the Internet, from the "root zone" [37, 38] through Top Level

Domains (TLDs, for example .org) to leaf domains (for example, example.org). *DNS Resolvers*, deployed most frequently within Internet Service Providers (ISPs) and lately by large cloud providers, query DNS authoritative servers searching for answers to queries sent to them by their client's Internet applications.

To reduce DNS service latency and to minimize outages, large DNS authoritative operators configure their authoritative DNS servers with various levels of redundancy: at the DNS level (by using multiple authoritative servers with different NS records[25]), at the IP level (by employing IP anycast [24, 32] so the same address is announced from multiple global locations/sites), and locally (by running multiple physical servers at each location/site, often deployed behind load balancers [27]).

On the resolver side, IP anycast is also used in some services, most typically in large-scale public ones (like the Quad{1,8,9} resolvers [1, 17, 34]). Caching within DNS resolvers is used to eliminate redundantly issued queries from being sent to authoritative servers, improving response times to users [28] and protecting users from short-term outages [29]. Caching also makes estimating the population size behind a resolver difficult, as authoritative servers only see queries on cache-misses.

In this work, we analyze queries sent from CP's resolvers towards authoritative name servers of two ccTLDs and B-Root. Given we only see DNS cache misses, our centralization analysis focuses on resolver to authoritative server DNS traffic, and not on user to authoritative.

2.1 Analyzed ccTLDs

We analyze authoritative DNS traffic (incoming queries) to .nl and .nz. From each ccTLD, we evaluate one week of data from the past three years, as can be seen in Table 2. To compensate for weekly diurnal patterns of the Internet [35], we choose yearly snapshots (similar to DNS-OARC DITL's dataset [14]) of one week's traffic, allowing us to compare annual changes. Table 2 also shows how many authoritative servers each ccTLD had at a time, and from how many we collected data, as well as the zone size at the time. To collect and analyze this data, both ccTLDs employed the same open-source DNS analysis platform (ENTRADA [40, 47]), allowing the same code to be run at each location.

2.1.1 .nl. We analyze traffic collected at two authoritative severs from .nl, out of a total of 4 (in 2018 and 2019) and 3 (in 2020), as seen in Table 2. Given these servers are deployed using anycast, they are actually distributed across a dozen global locations. These two authoritative servers are, in turn, run by two large independent DNS providers. We include only the three authoritative servers in this analysis that support pcap collection. We also show in Table 2 the size of the DNS zone distributed during each week. For .nl, all domains underneath are registered as second-level domains (as the second-level example.com is registered under .com).

2.1.2 .nz. We analyze traffic collected at 6 of the set of 7 authoritative servers for .nz (Table 2) during the collection periods. Similarly to .nl, we omit one of the authoritative servers from the analysis since it did not support pcap collection. Given .nz allows registrations as a third-level domain (for example example.net.nz) as well as a second-level domain (for example example.nz), we analyze traffic to .nz and all its second-level domains altogether. In total, the .nz and its subzones size ranged from 710 to 720k domain names. For the evaluated period, .nz had 140-141K second-level domains and 569-580K third-level domains.

Week	NSSet	Analyzed NSes	Zone size
		.nl	
w2018: Nov. 4–10, 2018	4A	2A	5.8M
w2019: Nov. 3–9, 2019	4A	2A	5.8M
w2020: April 5–11, 2020	3A	2A	5.9M
		.nz	
w2018: Nov. 4–10, 2018	6A, 1U	5A, 1U	720K
w2019: Nov. 3–9, 2019	6A, 1U	5A, 1U	710K
w2020: April 5–11, 2020	6A, 1U	5A, 1U	710K

Table 2: .nl and .nz authoritative servers. (A=Anycast, U=Unicast)

2.2 Analyzed Root Server Traffic

The DNS root servers sit at the top of the DNS infrastructure, and are contacted first by Internet resolvers while trying to find contact points for TLDs. They receive both legitimate requests, with query names that are either TLDs themselves or domain names underneath the TLDs, and illegitimate requests for query names without a real TLD as the suffix.

We analyze the data from samples of the Day in The Life of the Internet (DITL) [14] collections from B-Root, one of the 13 root servers. Because B-Root is deployed on multiple anycast sites around the world, we expect widely distributed query sources.

3 DATASETS

Table 3 shows the details of the datasets we analyze in this paper. In total, we study 55.7 billion DNS queries and responses (~30 billion for .nl, ~12 billion for .nz, and 14 billion for B-Root).

Traffic growth: we see that both ccTLDs and B-Root saw a traffic increase in the observed years. .nz, which kept the same number of authoritative servers in the period, saw a increase of 55% in query volume. .nl, in turn, saw an increase in 88% in the same period. However, part of this growth is also due to the fact that in the period, .nl went from 4 to 3 authoritative servers – and the extra queries were also captured by the two monitored authoritatives we analyzed data from. From 2018-2020, B-Root saw a significant increase in traffic (more than doubling – with a 150% increase). Some of this is due to the natural trending increase in DNS traffic to the Root server system. However, B-Root also increased its number of anycast sites, increasing its global footprint and attracting more traffic from additional nearby resolvers [30]. This resulted in a substantial growth in the number of resolvers (42%) and ASes (14 %) seen at B-Root.

"Junk" traffic: we define junk traffic as any query that does not yield a NOERROR RCODE (0 [26]). We see that the majority of queries are valid for .nl (~ 86%) and .nz (71%). The Root authoritative servers, however, experience a far larger volume of junk queries. We compute the query distribution for 11 out of the 13 Root Server that published aggregate statistics [21, 45] and found that only 32%, 23%, and 22% of queries were actually valid for w2018, w2019, and w2020, respectively. In the 2020/05/06 dataset used in this study, only 20% of B-Root's traffic consisted of valid queries, in sharp contrast with the TLDs. The Roots have been known since mid-2000's to receive high levels of junk traffic [10], but the traffic

.nl				
Week	**Queries(total)**	**Queries (valid)**	**Resolvers**	**ASes**
w2018	7.29B	6.53B	2.09M	41276
w2019	10.16B	9.05B	2.18M	42727
w2020	13.75B	11.88B	1.99M	41716
.nz				
Week	**Queries(total)**	**Queries (valid)**	**Resolvers**	**ASes**
w2018	2.95B	2.00B	1.28M	37623
w2019	3.48B	2.81B	1.42M	39601
w2020	4.57B	3.03B	1.31M	38505
B-Root				
Date	**Queries(total)**	**Queries (valid)**	**Resolvers**	**ASes**
2018/04/10	2.68B	0.93B	4.23M	45210
2019/04/09	4.13B	1.43B	4.13M	48154
2020/05/06	6.70B	1.34B	6.01M	51820

Table 3: Evaluated datasets.

has grown significantly since Chromium-based web browsers now intentionally generate random, non-existing TLD names (junk) during network initialization [19, 42].

4 CLOUD QUERIES

4.1 How much traffic comes from the clouds?

What percentage actually originates from CPs? Figure 1 shows the query distribution per CP and year, for *all* queries (total in Table 3).

5 CPs > 30% of ccTLD traffic. For .nl, we see that roughly 1/3 of all queries were from these CPs– with a slight growth from 2018 to 2019. For .nz, in turn, the concentration of traffic on the five providers is slightly smaller (less than 30% in 2019), but the same idea still holds: a significant concentration of DNS queries from only 20 ASes, out of the total of more than than 37,000 (Table 1) that sent queries in the monitored period.

B-Root, in turn, only received 8.7% of its traffic from CPs, as seen in Figure 1c. Root servers see queries from a wider view of the world, and are thus likely receiving more traffic from areas of the world where CPs may not have as much penetration. For example, in the 2020/04/06 dataset, the first CP was in a 5th place rank behind ISPs from India, France (2), and Indonesia. However, the growth in percentage of the CPs ASes in the last few years (Figure 1c) shows that wider penetration may be occurring at a slower rate.

Google issues more queries to .nl *than* .nz*.* The main difference in traffic between both ccTLDs comes from Google, which sends more in .nl traffic. Given that Google provides public DNS services, one hypothesis is that Google Public DNS would be more *popular* for .nl users than .nz's users. We can indirectly estimate this by determining the ratio of queries from Google's (Table 1) advertised list of IP addresses used by its Public DNS service [18].

Table 4 shows the classification of queries from Google's Public DNS verses the remaining part of its infrastructure (for example, from its internal corporate DNS, or the DNS services used on within their cloud services) for the w2020 dataset (Table 3). In both cases, Google Public DNS is responsible for about the same ratio of queries from Google – 86.5% and 88.4% (w2019 has similar values, as seen in Appendix A). Given that both countries have similar ratios of Public DNS queries from Google, this does not explain why Google

	.nl		.nz	
	Queries	Resolv.	Queries	Resolv.
Total	1.81B	23943	328.7M	21230
Pub. DNS	1.57B	3750	290.7M	3840
Rest	0.24B	23943	38.0M	17390
Ratio Pub.	86.5%	15.6%	88.4%	18.7%

Table 4: Queries from Google on w2020

sends more queries to .nl than .nz. As such, it may be simply a difference of Google market penetration in both countries.

4.2 What records do clouds ask for?

DNS stores multiple types of *resource records* (RRs) [26]. A and AAAA records map domain names to IPv4 and IPv6 addresses, respectively. NS records, in turn, store names of a domain's authoritative servers. MX records indicate where to find a domain's e-mail servers, while DS and DNSKEY records are used in DNSSEC [3–5].

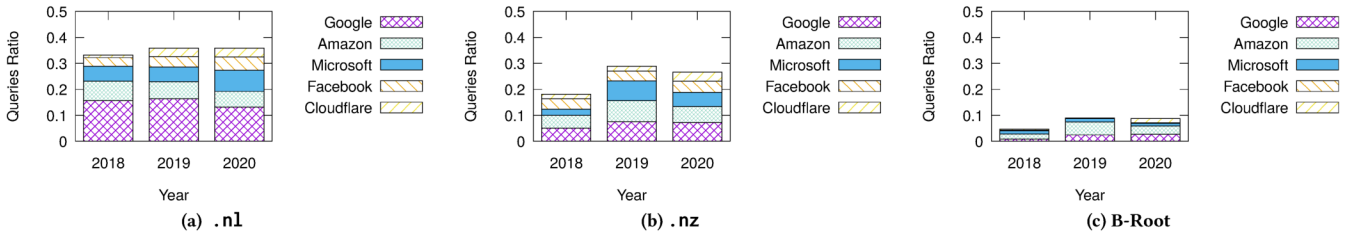
To see if cloud resolvers have different end-goals that manifest in different resolution behaviors, we examine the types of requests they send. Distinct differences are visible in Figure 2 between 2018 and 2020 (we omit 2019 for space reasons, and include it in Appendix B). Comparing the ccTLDs to each other, we see both similarities and differences in CP requests. The similarities are highlighted in the distribution of requested RRs. For example, in 2018, the most popular record type was A for both ccTLDs and for B-Root. The rest of this section discusses the observed differences in cloud queries.

4.2.1 Qname minimization deployment. Figure 2 shows that in 2018 for the ccTLDs that most query types were for A and AAAA records. However, in 2020 the number of NS queries dramatically increased for 3 of the 5 CPs, for both ccTLDs (and Amazon for .nz).

One root cause may be the deployment of QNAME minimization (Q-min hereafter), a technique to improve privacy by sending only the necessary data in requests to authoritative servers. To do this, resolvers query servers with query names "stripped to just one label more than the zone for which the server is authoritative" [9], minimizing the extra label names "leaked" to the server.

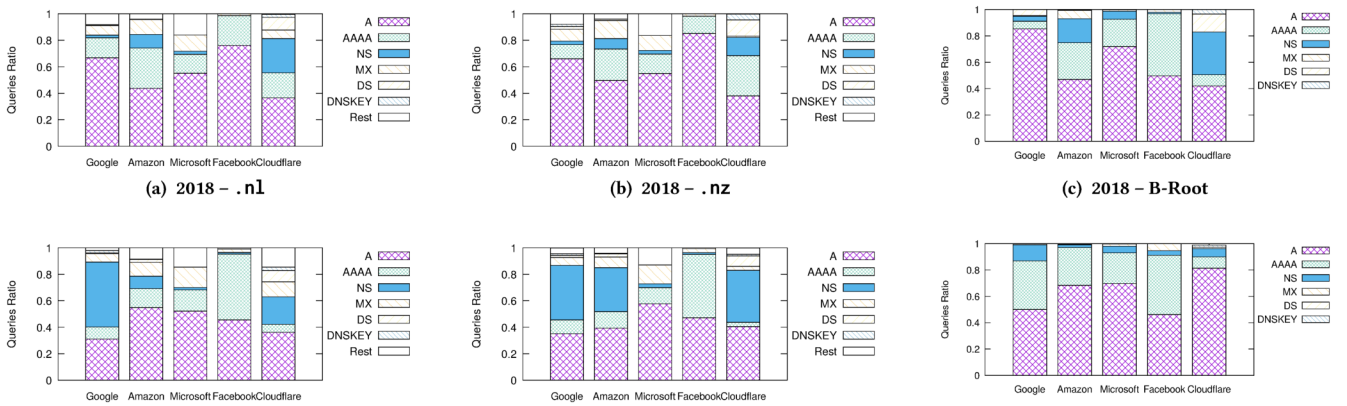
Q-min requires resolvers to first query for the NS records of a requested domain, and subsequently query those servers for additional information. The increase in the percentage of queried NS records may indicate the deployment of Q-min within the CPs. Analyzing the queries from each CP confirmed our suspicions: CPs with a significant growth of NS queries queried for minimized names.

Through a finer grained longitudinal study, we can actually *determine* when Q-min was adopted by each provider. To illustrate this, we show monthly queries from Google to the ccTLDs in Figure 3. In Dec. 2019 we see the first increase in the number of NS queries, for both ccTLDs. After manually verifying the query names to ensure they match expected Q-min behavior, we reached out to Google operators, who confirmed that Q-min deployment did take place in Dec. 2019. We see that after that , the proportion of NS queries remains high for both ccTLDs from Jan–April 2020 (the exception is in Figure 3b, where Google sends more A/AAAA queries in Feb2020 for .nz. The causes for that was a cyclic dependency [31], a type of DNS misconfiguration on two .nz domain names that took place in Feb 2020, and caused Google to issue millions of A/AAAA queries.



(a) `.nl` (b) `.nz` (c) B-Root

Figure 1: Clouds query ratio per ccTLD and B-Root



(a) 2018 – `.nl` (b) 2018 – `.nz` (c) 2018 – B-Root

(d) 2020 – `.nl` (e) 2020 – `.nz` (f) 2020 – B-Root

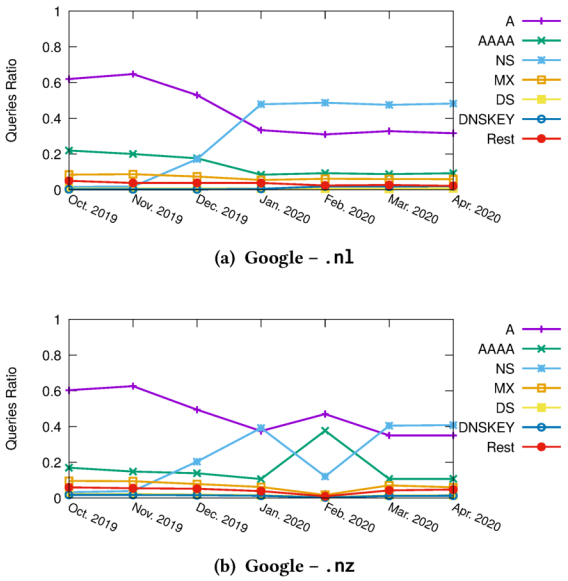
Figure 2: Resource Records per Cloud provider

Once this misconfiguration was solved, the trend of higher number of NS queries resumes (March and April 2020).

Overall, the growth in NS queries for both ccTLDs and the `Q-min` deployment at a large CP such as Google shows actually a positive side of consolidation: once a large player adopt a new technology that enhances privacy such as `Q-min`, it ends up improving the privacy for all of its users. In contrast, a negative deployment aspect would have just as rapid a roll out of problems to its users as well.

4.2.2 DNSSEC validation deployment. DNSSEC validation refers to resolvers verifying their received answers with cryptographically signed answers from the authoritative server. To perform such operations, resolvers need to retrieve DS and DNSKEY records.

To some extent, all CPs seem to validate DNS queries, *except for one*, as can be seen in Figure 2. This is to show that there is a discrepancy in the adoption of more secure technologies in CPs. From an authoritative source, a resolver should only query for DNSKEYs once in a TTL length. However, it will likely query for DS records for all of the sub-domains of an authoritative source. This can be visually seen in Figure 2d, where Cloudflare (a known DNSSEC validating resolver) sends more queries for DS records than DNSKEY records. Interestingly, Google's widely-used public DNS service is also a validating resolver, but the percentage of queries for DS records is fairly low in all of the Figure 2 graphs. This may be because the queries from the rest of the Google infrastructure dwarfs the DNSSEC validation related queries from their public



(a) Google – `.nl`

(b) Google – `.nz`

Figure 3: Queries distribution per month for Google.

		.nl				.nz			
	Year	IPv4	IPv6	UDP	TCP	IPv4	IPv6	UDP	TCP
Google	2018	0.66	0.34	1	0	0.61	0.39	1	0
	2019	0.49	0.51	1	0	0.54	0.46	1	0
	2020	0.52	0.48	1	0	054	0.46	1	0
Amazon	2018	1	0	1	0	1	0	0.98	0.02
	2019	0.98	0.02	0.98	0.02	0.97	0.03	0.96	0.04
	2020	0.97	0.03	0.95	0.05	0.96	0.04	0.95	0.05
Microsoft	2018	1	0	1	0	1	0	1	0
	2019	1	0	1	0	1	0	1	0
	2020	1	0	1	0	1	0	1	0
Facebook	2018	0.52	0.48	0.79	0.21	0.51	0.49	0.52	0.48
	2019	0.24	0.76	0.85	0.15	0.19	0.81	0.83	0.17
	2020	0.24	0.76	0.86	0.14	0.17	0.83	0.85	0.15
Cloudflare	2018	0.54	0.46	1	0	0.54	0.46	1	0
	2019	0.57	0.43	0.99	0.01	0.56	0.44	1	0
	2020	0.51	0.49	0.98	0.02	0.49	0.51	0.99	0.01

Table 5: Query Distribution per CP for ccTLDs

resolver. For example, in w2020, Google has sent ~10M DS queries to .nl, out of the 1.8B of all queries it sent.

4.2.3 DNS "junk" per cloud. In Figure 4 we show the proportion of "junk" (non RCODE 0) queries compared to all the queries each CP sends. We see that .nl and .nz have similar junk rates. B-Root, which receives 80% junk queries overall (§3), sees proportionally *fewer* junk queries from CPs (except for Cloudflare in 2019). It is possible that the decrease in junk queries from CP's in 2020 corresponds with potential deployments of NSEC aggressive caching [16].

4.3 IPv4 vs IPv6 usage

We now turn to the resolver infrastructure of each CP. Table 5 summarizes the query distribution for each CP in terms of network and transport protocol. We see a large variation in IPv4 vs IPv6 usage: Amazon and Microsoft send roughly all its queries over IPv4. This is quite surprising for such large providers – one might expect that CPs would be among the early adopters of IPv6 (all authoritative name servers analyzed in this study offer both IPv4 and IPv6 service). Table 5 shows that Cloudflare and Google use IPv4 and IPv6 more evenly. And Facebook, since 2019, actually sends more queries over IPv6 than IPv4.

Why do Amazon and Microsoft make little use of IPv6? Comparing Table 6 to Table 5, we see a direct correlation between the number of resolvers per IP version and traffic: for .nl, only 1.8% of the Amazon's resolvers are IPv6, and they send 3% of the queries (w2020). For Microsoft, we see 3% of resolvers, but the traffic is much smaller. We see similar figures for .nz: Amazon with 2.1% of resolvers being IPv6, while Microsoft having 4.6% of resolvers.

Why does Facebook favors IPv6? to understand why Facebook prefers IPv6 over IPv4, we hypothesize that IPv6 queries are answered faster than IPv4, given that resolvers *tend* to favor authoritative servers with lower latency [30].

To determine if IPv6 preference stems from latency, we first determine which resolvers are dual-stack, and calculate their round-trip times (RTTs) for each IP version. We determine which resolvers are dual-stack by using reverse DNS [22] to "reverse lookup" their IP addresses and turn it back into a domain name. We perform this process for each IP address that sent queries to Facebook.

	.nl	.nz
Amazon	38317	34645
IPv4	37640 (98.2%)	33908 (97.9%)
IPv6	677 (1.8%)	737 (2.1%)
Microsoft	14494	10206
IPv4	14069 (97.0%)	9738 (95.4%)
IPv6	425 (3.0%)	468 (4.6%)

Table 6: Amazon and Microsoft resolvers (Week 2020)

Facebook, like many operational DNS services, includes site locations within the PTR records names returned by the reverse lookup process. We identify 13 different sites based on the airport codes embedded in the returned PTR names. For 12 of these sites, the PTR record names also include the IPv4 address of the host – even if the queried IP address is IPv6. By using reverse lookup of all received IPv4 and IPv6 addresses, we can identify IP address who's multiple PTR records refer to the same resolver, thus categorizing them as a *dual-stack* resolver. Only 1 IPv4 and 2 IPv6 addresses had no PTR record associated with it (2020-04-20).

For each dual-stack identified resolver, we single out the TCP queries in 2020 (14% of total, as shown in Table 5) and calculate the *median* RTT of the TCP handshakes.

Figure 5a shows each of Facebook's resolvers location and its respective query distribution over IPv4 and IPv6. We see location 1 dominates the query volumes in comparision to other locations. Locations 8–10 send more queires over IPv4, and the remaning ones have a more evenly distribution of queries.

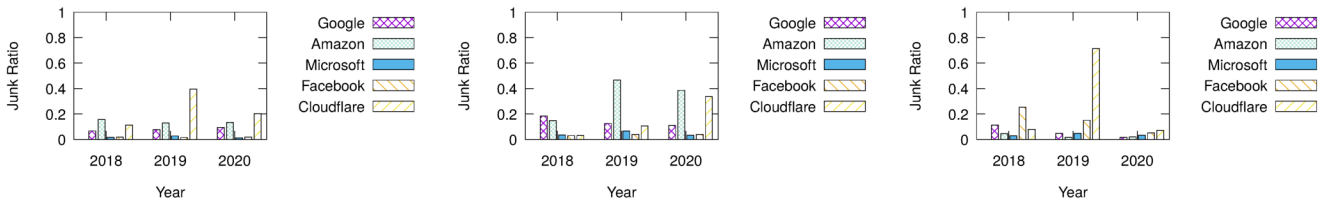
Figure 5b shows the IPv6 query ratio for Facebook, *per location*, and their respective median RTTs, for queries sent to Server A of .nl (we include Server's B graph in Appendix B). For Location 1, we observed no TCP traffic, so we cannot estimate its RTT (this also shows how the query behavior within a cloud is not homogeneous). Thus, we were unable to verify our hypothesis that Facebook optimizes RTT times by favoring IPv6 RTT since their dominating location does not send any DNS queries over TCP.

However, we show that for other locations, the RTT correlates with whether queries are sent over IPv4 or IPv6. We see that locations 8–10, which have significantly larger IPv6 RTTs, prefer IPv4 over IPv6. Server B shows a similar behavior: locations 2 and 4 receive more IPv4 queries given their larger RTT differences – Appendix B. This behavior confirms the findings of a previous study [30]. The remaining locations have similar IPv4 and IPv6 RTTs, explaining their more even query distribution.

4.4 UDP vs TCP

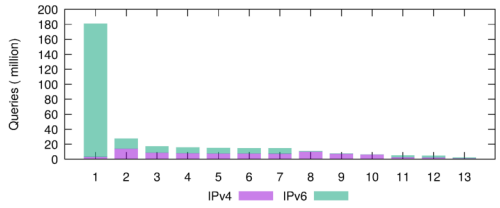
Except for Facebook, all CPs send very few DNS queries over TCP, as seen in Table 5. TCP is rarely needed by DNS protocol implementations, except when transferring larger data sets (such as retrieving DNSKEYs during DNSSEC validation), or when resolvers hit a "Response Rate Limiting" [44] threshold, which requires them to switch to TCP to prove they are not spoofing UDP requests.

Traditional DNS messages are limited to 512 bytes when using UDP [26]. Extension Mechanisms for DNS (EDNS0) [11, 43] enabled resolvers to inform authoritative servers if they can handle larger UDP messages (UDP Message Size). If the advertised size is smaller

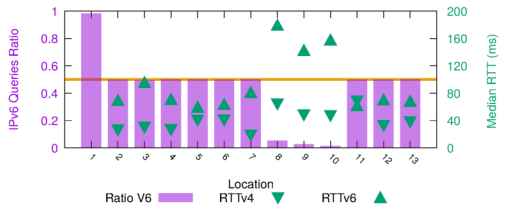


(a) `.nl` (b) `.nz` (c) B-Root

Figure 4: Clouds' DNS junk query ratio per ccTLD and B-Root



(a) Facebook Location vs Queries to `.nl`'s Server A (w2020).



(b) Ratio queries IPv6 and RTT to Server A of `.nl` in w2020.

Figure 5: Facebook Resolver's location and IPv4 and IPv6 usage when querying `.nl`'s Server A (w2020) .

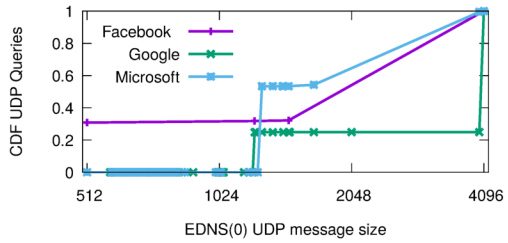


Figure 6: CDF of EDNS(0) UDP message size for `.nl` (w2020).

than the answer, the authoritative server will truncate the answer. The resolver, in turn, should query again using TCP instead.

To investigate if EDNS(0) UDP message size affects TCP usage, we show in Figure 6 the CDF of Facebook and Google's UDP message sizes for `.nl`. We see that roughly 30% of Facebook UDP queries had a EDNS(0) message size of 512 bytes, which is very different from Google, which 24% of queries have EDNS(0) sizes up to 1232 bytes (Microsoft has similar figures to Google). This, in turn, may explain the ratio of truncated UDP answers: for w2020, Facebook had 17.16% UDP answers truncated, while Google had 0.04% and Microsoft had 0.01%. These different usage of transport protocol, again, shows the variation of CP's resolver infrastructure.

5 RELATED WORK

Our work is the first to compare CPs against each other in terms of DNS traffic to two ccTLDs and one Root Server letter. A previous work [12] analyzed traffic from Google Public DNS towards the authoritative servers of various zones. We instead focus on 5 CPs – including Google's public DNS and *its remaining* infrastructure.

`Q-min` has been analyzed using *active measurements* using both RIPE Atlas probing and data from passive authoritative name servers from `.nl` and K-root [13]. It authors showed that 33-40% of queries were minimized. However, they did not break down the results per CP. Our results show, for example, exactly *when* Google deployed `Q-min` in its Public Service. Radu *et al.* [36] analyzes DNS traffic centralization by measuring mobile devices using OONI, a tool developed by the TorProject to detect network anomalies and censorship. As the authors acknowledge, their datasets are skewed towards tech savvy users, mostly from mobile devices, and it is not representative for all Internet users. We consider traffic from the authoritative side, between resolvers and authoritative servers, which rarely comes directly from users. Last, Foremski *et al.* [15] analyzed traffic from passive data from various resolvers and found a very small fraction of `Q-min` queries.

With regards DNS centralization, Allman [2] covers infrastructure centralization – which does not account how popular domains in a zone are. By analyzing DNS traffic, we look into actual resource usage. Yeganeh *et al.* [48] compares CPs in terms of performance and connectivity – we focus on DNS traffic distribution.

6 CONCLUSIONS

We measure Internet centralization by analyzing DNS traffic and show that 5 large CPs, from their 20 ASes, are responsible for 30% of queries to two ccTLDs but only 8.7% of B-Root, over the last three years. This is remarkable concentration considering we observed traffic from 37k+ other ASes to the ccTLDs and root server.

Like real clouds, CPs are not all alike: we show a large variation in terms of technology adoption – some completely ignore IPv6 and DNSSEC altogether, while others actively prefer IPv6, and choosing END(0) parameters, impacting TPC usage. An observed positive side of centralization to CPs is the rapid deployment of a secure/privacy feature, covering many users, seen when Google deployed `Q-min` in Dec. 2019.

Acknowledgments

We thank the anonymous IMC reviewers for their time and our shepherd, Mark Allman.

The work of Giovane C. M. Moura, Maarten Wullink and Cristian Hesselman is partially funded by the European Union's Horizon 2020 CONCORDIA project (Grant Agreement 830927) and by the Plannning for Anycast as Anti-DDoS research project (PAADDoS), under the research programme Cyber Security (project number 628.001.029), which is funded by the Dutch Research Council (NWO).

Wes Hardaker's work in this paper is partially supported by USC as part of B-Root research activity.

REFERENCES

[1] 1.1.1.1. 2018. The Internet's Fastest, Privacy-First DNS Resolver. https://1.1.1.1/. https://1.1.1.1/

[2] Mark Allman. 2018. Comments on DNS Robustness. In *Proceedings of the Internet Measurement Conference 2018* (Boston, MA, USA) *(IMC '18)*. Association for Computing Machinery, New York, NY, USA, 84–90. https://doi.org/10.1145/3278532.3278541

[3] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. 2005. *DNS Security Introduction and Requirements*. RFC 4033. IETF. http://tools.ietf.org/rfc/rfc4033.txt

[4] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. 2005. *Protocol Modifications for the DNS Security Extensions*. RFC 4035. IETF. http://tools.ietf.org/rfc/rfc4035.txt

[5] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. 2005. *Resource Records for the DNS Security Extensions*. RFC 4034. IETF. http://tools.ietf.org/rfc/rfc4034.txt

[6] J. Arkko. 2019. Centralised Architectures in Internet Infrastructure. Internet Draft. https://tools.ietf.org/html/draft-arkko-arch-infrastructure-centralisation-00

[7] Jari Arkko. 2020. The influence of Internet architecture on centralised versus distributed Internet services. *Journal of Cyber Policy* 5, 1 (2020), 30–45. https://doi.org/10.1080/23738871.2020.1740753

[8] Arkko, Jari and Tramme, B. and Nottingham, M and Huitema, C and Thomson, M. and Tantsura, J. and ten Oever, N. 2019. Considerations on Internet Consolidation and the Internet Architecture. Internet Draft. https://tools.ietf.org/html/draft-arkko-iab-internet-consolidation-02

[9] S. Bortzmeyer. 2016. *DNS Query Name Minimisation to Improve Privacy*. RFC 7816. IETF. http://tools.ietf.org/rfc/rfc7816.txt

[10] Sebastian Castro, Duane Wessels, Marina Fomenkov, and Kimberly Claffy. 2008. A Day at the Root of the Internet. *ACM Computer Communication Review* 38, 5 (April 2008), 41–46.

[11] J. Damas, M. Graff, and P. Vixie. 2013. *Extension Mechanisms for DNS (EDNS(0))*. RFC 6891. IETF. http://tools.ietf.org/rfc/rfc6891.txt

[12] Wouter B. De Vries, Roland Van Rijswijk-Deij, Pieter Tjerk De Boer, and Aiko Pras. 2018. Passive Observations of a Large DNS Service: 2.5 Years in the Life of Google. In *2018 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, United States. https://doi.org/10.23919/TMA.2018.8506536

[13] Wouter B de Vries, Quirin Scheitle, Moritz Müller, Willem Toorop, Ralph Dolmans, and Roland van Rijswijk-Deij. 2019. A First Look at QNAME Minimization in the Domain Name System. In *International Conference on Passive and Active Network Measurement*. Springer, 147–160.

[14] DNS OARC. 2020. DITL Traces and Analysis. https://www.dns-oarc.net/index.php/oarc/data/ditl/.

[15] Pawel Foremski, Oliver Gasser, and Giovane C. M. Moura. 2019. DNS Observatory: The Big Picture of the DNS. In *Proceedings of the Internet Measurement Conference* (Amsterdam, Netherlands) *(IMC '19)*. Association for Computing Machinery, New York, NY, USA, 87–100. https://doi.org/10.1145/3355369.3355566

[16] K. Fujiwara, A. Kato, and W. Kumari. 2017. *Aggressive Use of DNSSEC-Validated Cache*. RFC 8198. IETF. http://tools.ietf.org/rfc/rfc8198.txt

[17] Google. 2019. Google Public DNS. https://developers.google.com/speed/public-dns/

[18] Google. 2020. Google Public DNS: Frequently Asked Questions. https://developers.google.com/speed/public-dns/faq

[19] Wes Hardaker. [n.d.]. What's in a name? https://blog.apnic.net/2020/04/13/whats-in-a-name/ https://blog.apnic.net/2020/04/13/whats-in-a-name/.

[20] P. Hoffman, A. Sullivan, and K. Fujiwara. 2018. *DNS Terminology*. RFC 8499. IETF. http://tools.ietf.org/rfc/rfc8499.txt

[21] ICANN. 2014. RSSAC002: RSSAC Advisory on Measurements of the Root Server System. https://www.icann.org/en/system/files/files/rssac-002-measurements-root-20nov14-en.pdf.

[22] P. Johansson. 1999. *IPv4 over IEEE 1394*. RFC 2734. IETF. http://tools.ietf.org/rfc/rfc2734.txt

[23] Cecilia Kang and David McCabe. 2020. Lawmakers, United in Their Ire, Lash Out at Big Tech's Leaders. *New York Times* (July. 29 2020). https://www.nytimes.com/2020/07/29/technology/big-tech-hearing-apple-amazon-facebook-google.html

[24] D. McPherson, D. Oran, D. Thaler, and E. Osterweil. 2014. *Architectural Considerations of IP Anycast*. RFC 7094. IETF. http://tools.ietf.org/rfc/rfc7094.txt

[25] P.V. Mockapetris. 1987. *Domain names - concepts and facilities*. RFC 1034. IETF. http://tools.ietf.org/rfc/rfc1034.txt

[26] P.V. Mockapetris. 1987. *Domain names - implementation and specification*. RFC 1035. IETF. http://tools.ietf.org/rfc/rfc1035.txt

[27] Giovane C. M. Moura, Ricardo de O. Schmidt, John Heidemann, Wouter B. de Vries, Moritz Müller, Lan Wei, and Christian Hesselman. 2016. Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event. In *Proceedings of the ACM Internet Measurement Conference* (johnh: pafile). ACM, Santa Monica, California, USA, 255–270. https://doi.org/10.1145/2987443.2987446

[28] Giovane C. M. Moura, John Heidemann, Ricardo de O. Schmidt, and Wes Hardaker. 2019. Cache Me If You Can: Effects of DNS Time-to-Live. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Amsterdam, the Netherlands, 101–115. https://doi.org/10.1145/3355369.3355568

[29] Giovane C. M. Moura, John Heidemann, Moritz Müller, Ricardo de O. Schmidt, and Marco Davids. 2018. When the Dike Breaks: Dissecting DNS Defenses During DDoS. In *Proceedings of the ACM Internet Measurement Conference* (johnh: pafile). Boston, MA, USA, 8–21. https://doi.org/10.1145/3278532.3278534

[30] Moritz Müller, Giovane C. M. Moura, Ricardo de O. Schmidt, and John Heidemann. 2017. Recursives in the Wild: Engineering Authoritative DNS Servers. In *Proceedings of the ACM Internet Measurement Conference*. ACM, London, UK, 489–495. https://doi.org/10.1145/3131365.3131366

[31] Vasileios Pappas, Zhiguo Xu, Songwu Lu, Daniel Massey, Andreas Terzis, and Lixia Zhang. 2004. Impact of Configuration Errors on DNS Robustness. *SIGCOMM Comput. Commun. Rev.* 34, 4 (Aug. 2004), 319–330. https://doi.org/10.1145/1030194.1015503

[32] C. Partridge, T. Mendez, and W. Milliken. 1993. *Host Anycasting Service*. RFC 1546. IETF. http://tools.ietf.org/rfc/rfc1546.txt

[33] Nicole Perlroth. 2016. Hackers Used New Weapons to Disrupt Major Websites Across U.S. *New York Times* (Oct. 22 2016), A1. http://www.nytimes.com/2016/10/22/business/internet-problems-attack.html

[34] Quad9. 2018. Quad9 | Internet Security & Privacy In a Few Easy Steps. https://quad9.net.

[35] Lin Quan, John Heidemann, and Yuri Pradkin. 2014. When the Internet Sleeps: Correlating Diurnal Networks with External Factors. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. ACM, New York, NY, USA, 87–100. https://doi.org/10.1145/2663716.2663721

[36] Roxana Radu and Michael Hausding. 2020. Consolidation in the DNS resolver market – how much, how fast, how dangerous? *Journal of Cyber Policy* 5, 1 (2020), 46–64. https://doi.org/10.1080/23738871.2020.1722191

[37] Root Server Operators. 2020. Root DNS. http://root-servers.org/.

[38] Root Zone file. 2020. Root. http://www.internic.net/domain/root.zone.

[39] Bruce Schneier. 2018. Censorship in the Age of Large Cloud Providers. https://www.schneier.com/essays/archives/2018/06/censorship_in_the_ag.html

[40] SIDN Labs. 2020. ENTRADA - DNS Big Data Analytics. https://entrada.sidnlabs.nl/.

[41] Internet Society. 2019. Consolidation in the Internet Economy. https://future.internetsociety.org/2019/

[42] Matthew Thomas. [n.d.]. Chromium's impact on root DNS traffic. https://blog.apnic.net/2020/08/21/chromiums-impact-on-root-dns-traffic/ https://blog.apnic.net/2020/08/21/chromiums-impact-on-root-dns-traffic/.

[43] P. Vixie. 1999. *Extension Mechanisms for DNS (EDNS0)*. RFC 2671. IETF. http://tools.ietf.org/rfc/rfc2671.txt

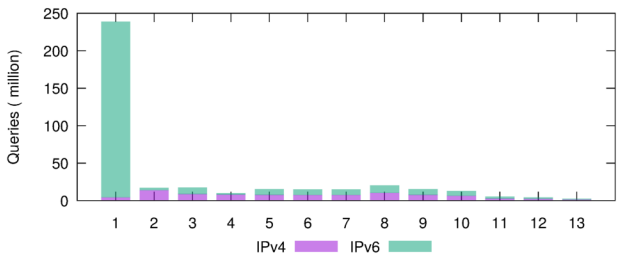
[44] Paul Vixie. 2014. Rate-Limiting State. *Commun. ACM* 57, 4 (April 2014), 40–43. https://doi.org/10.1145/2578902

[45] Duane Wessels. 2020. RSSAC002-data. https://github.com/rssac-caucus/RSSAC002-data/.

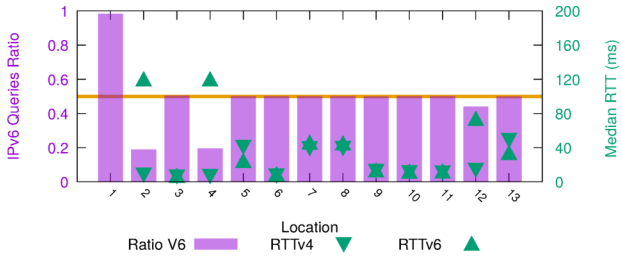
[46] Chris Williams. 2019. Bezos DDoS'd: Amazon Web Services' DNS systems knackered by hours-long cyber-attack. https://www.theregister.co.uk/2019/10/22/aws_dns_ddos/.

[47] Maarten Wullink, Giovane CM Moura, Moritz Müller, and Cristian Hesselman. 2016. ENTRADA: A high-performance network traffic data streaming warehouse. In *Network Operations and Management Symposium (NOMS), 2016 IEEE/IFIP*. IEEE, 913–918.

[48] Bahador Yeganeh, Ramakrishnan Durairajan, Reza Rejaie, and Walter Willinger. 2020. A First Comparative Characterization of Multi-cloud Connectivity in Today's Internet. In *International Conference on Passive and Active Network Measurement*. Springer, 193–210.



(a) Facebook Location vs Queries to `.nl`'s Server B (w2020).

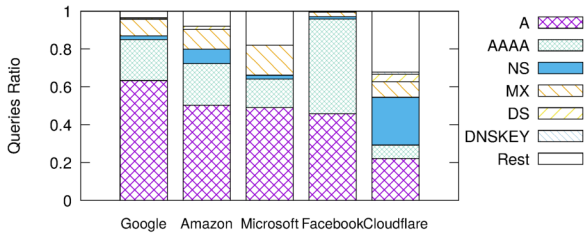


(b) Ratio queries IPv6 and RTT to Server B of `.nl` in w2020.

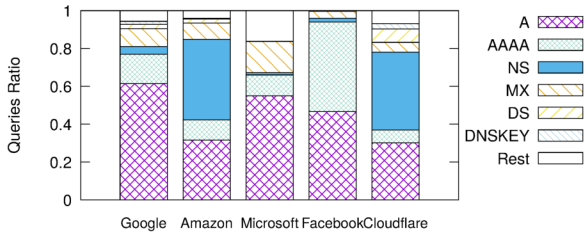
Figure 8: Facebook Resolver's location and IPv4 and IPv6 usage when querying `.nl`'s Server B (w2020) .

	`.nl`		`.nz`	
	Queries	Resolv.	Queries	Resolv.
Total	1.6B	23344	263.8M	20089
Pub. DNS	1.49B	3581	222.9M	3575
Rest	0.176B	19754	40.1M	16514
Ratio Pub.	89.3%	15.4%	84.4%	17.7%

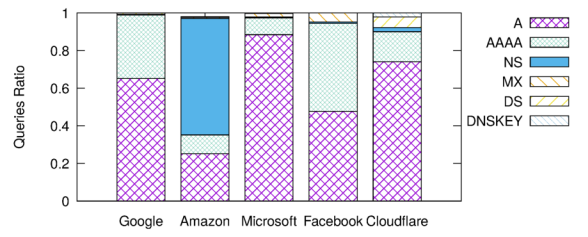
Table 7: Queries from Google on w2019



(a) 2019 – `.nl`



(b) 2019 – `.nz`



(c) 2019 – B-Root

Figure 7: Resource Records per Cloud provider for 2019

A EXTRA TABLES

Table 7 show the query distribution for Google in w2019, for its Public DNS service and other parts of the cloud.

B EXTRA GRAPHS

Figure 7 shows the extra graph for RR distribution for 2019.

Figure 8 shows Facebook query sources towards Server B, one of the other authoritative servers of `.nl`.