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

Giovane C. M. Moura (1), Sebastian Castro (2), Wes Hardaker (3), Maarten Wullink (1) and Cristian Hesselman (1,4)

DNS-OARC
2020-09-28
Virtual Conference

1: SIDN Labs 2: InternetNZ 3: USC/ISI 4: University of Twente
Internet centralization concerns: EU

‘This Is a New Phase’: Europe Shifts Tactics to Limit Tech’s Power

The region’s lawmakers and regulators are taking direct aim at Amazon, Facebook, Google and Apple in a series of proposed laws.

Internet centralization concerns: US DOJ

Justice Department Opens Antitrust Review of Big Tech Companies

Abstract

Centralised deployment models for Internet services and Internet business consolidation are well-known Internet trends, at least when it comes to popular and user-visible service. This memo discusses the impacts of similar trends within the Internet infrastructure, on functions such as DNS resolution.
Centralization poses various risks

- Creates a **single point of failure**
- Privacy
- Market consolidation

Centralization poses various risks

- Creates a single point of failure
- Privacy
- Market consolidation

Amazon **Route 53 (DNS)** 2019 Attack

Can we measure Internet Centralization?

Easier said than done.

Measure it in terms of?

- Users?
- Traffic?
- Networking infrastructure?
- Computing infrastructure?
- Market?
- ...
Can we measure Internet Centralization?

Easier said than done.

Measure it in terms of?

- Users?
- Traffic?
- Networking infrastructure?
- Computing infrastructure?
- Market?
- ...

Our approach:

- We focus on **DNS traffic**
- But **NOT** on *user* traffic
- We focus on traffic from resolvers to authoritative servers
What we measure: DNS queries to

The Netherlands (.nl)
- 17.1M inhabitants
- 6M domain names (.nl)
- Continent: Europe
- Official language: Dutch

New Zealand (.nz)
- 4.8 M inhabitants
- 700k domain names (.nz)
- Continent: Oceania
- Official languages: English, Maori

B-Root

World
- 7.8 Billion inhabitants
- 1588 TLDs
- Continents: 7
- Language: *

7
### What we measure: DNS queries from

**From 5 Cloud/Content Providers**

<table>
<thead>
<tr>
<th>Company</th>
<th>ASes</th>
<th>Public DNS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>15169</td>
<td>Yes</td>
</tr>
<tr>
<td>Amazon</td>
<td>7224, 8987, 9059, 14168, 16509</td>
<td>No</td>
</tr>
<tr>
<td>Microsoft</td>
<td>3598, 6584, 8068–8075, 12076, 23468</td>
<td>No</td>
</tr>
<tr>
<td>Facebook</td>
<td>32934</td>
<td>No</td>
</tr>
<tr>
<td>Cloudflare</td>
<td>13335</td>
<td>Yes</td>
</tr>
</tbody>
</table>
## Evaluated datasets.

<table>
<thead>
<tr>
<th>Week</th>
<th>Queries (total)</th>
<th>Queries (valid)</th>
<th>Resolvers</th>
<th>ASes</th>
</tr>
</thead>
<tbody>
<tr>
<td>w2018</td>
<td>7.29B</td>
<td>6.53B</td>
<td>2.09M</td>
<td>41276</td>
</tr>
<tr>
<td>w2019</td>
<td>10.16B</td>
<td>9.05B</td>
<td>2.18M</td>
<td>42727</td>
</tr>
<tr>
<td>w2020</td>
<td>13.75B</td>
<td>11.88B</td>
<td>1.99M</td>
<td>41716</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week</th>
<th>Queries (total)</th>
<th>Queries (valid)</th>
<th>Resolvers</th>
<th>ASes</th>
</tr>
</thead>
<tbody>
<tr>
<td>w2018</td>
<td>2.95B</td>
<td>2.00B</td>
<td>1.28M</td>
<td>37623</td>
</tr>
<tr>
<td>w2019</td>
<td>3.48B</td>
<td>2.81B</td>
<td>1.42M</td>
<td>39601</td>
</tr>
<tr>
<td>w2020</td>
<td>4.57B</td>
<td>3.03B</td>
<td>1.31M</td>
<td>38505</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Queries (total)</th>
<th>Queries (valid)</th>
<th>Resolvers</th>
<th>ASes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018/04/10</td>
<td>2.68B</td>
<td>0.93B</td>
<td>4.23M</td>
<td>45210</td>
</tr>
<tr>
<td>2019/04/09</td>
<td>4.13B</td>
<td>1.43B</td>
<td>4.13M</td>
<td>48154</td>
</tr>
<tr>
<td>2020/05/06</td>
<td>6.70B</td>
<td>1.34B</td>
<td>6.01M</td>
<td>51820</td>
</tr>
</tbody>
</table>

Table 1: Evaluated datasets.
So, what did we find?
Traffic to b.root-servers.net

<table>
<thead>
<tr>
<th>Year</th>
<th>Queries Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
</tr>
</tbody>
</table>

- **Google**
- **Amazon**
- **Microsoft**
- **Facebook**
- **Cloudflare**
The 5 clouds account for roughly 1/3 of all queries to .nl and .nz

- .nl and .nz see 40k+ Autonomous Systems in total
- b.root-servers.net receives less, with than 9% of traffic from clouds
  - likely affected by tons of chromium-based garbage [5, 6]
- Oddity: Google sends more traffic to .nl than .nz
What do clouds dream of when visiting the Netherlands?

(d) 2018 – .nl

(e) 2019 – .nl

(f) 2020 – .nl
What do clouds dream of when visiting New Zealand?

(g) 2018 – .nz

(h) 2019 – .nz

(i) 2020 – .nz
What do clouds dream of when visiting the Root?

(j) 2018 – B

(k) 2019 – B

(l) 2020 – B
What did clouds dream of in 2018?

(m) 2018 – .nl
(n) 2018 – .nz
(o) 2018 – B
What do clouds dream of in 2019?

(p) 2019 – .nl

(q) 2019 – .nz

(r) 2019 – B
What do clouds dream of in 2020?

(s) 2020 – .nl

(t) 2020 – .nz

(u) 2020 – B
What do clouds dream of?

Resource Records per Cloud provider
Mostly A records, but...
What do they ask for?

- Google sends more NS queries in 2020 than in 2018
- Why?
  - QNAME-minimization [4]
  - Q-min first query for the NS records
- We confirmed with Google that they deployed QNAME-minimization in Dec. 2019
Identifying when Google deployed RFC7816

- As Google deployed QNAME-minimization it created a visible shift in query types
- **Centralization Pro:** new security feature deployments benefits many users all at once
  - DNSSEC validation
  - QNAME-minimization

Queries distribution per month for Google.

(ad) Google – .nl
Junk queries sent to .NL from clouds

<table>
<thead>
<tr>
<th>Year</th>
<th>Junk Ratio</th>
<th>Cloud</th>
<th>Google</th>
<th>Amazon</th>
<th>Microsoft</th>
<th>Facebook</th>
<th>Cloudflare</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>0.000</td>
<td>0.200</td>
<td>0.400</td>
<td>0.600</td>
<td>0.800</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>0.000</td>
<td>0.200</td>
<td>0.400</td>
<td>0.600</td>
<td>0.800</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>0.000</td>
<td>0.200</td>
<td>0.400</td>
<td>0.600</td>
<td>0.800</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>
Junk queries sent to .NZ from clouds

![Graph showing junk queries sent to .NZ from clouds for Google, Amazon, Microsoft, Facebook, and Cloudflare in 2018, 2019, and 2020.](image-url)
Junk queries sent to b.root-servers.net from clouds

![Graph showing junk ratio for different cloud providers over years 2018 to 2020.]

- Google
- Amazon
- Microsoft
- Facebook
- Cloudflare

Junk Ratio (

2018 2019 2020

Year

(ag) b.root-servers.net
Junk queries raining from the clouds

<table>
<thead>
<tr>
<th>Year</th>
<th>Google</th>
<th>Amazon</th>
<th>Microsoft</th>
<th>Facebook</th>
<th>Cloudflare</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Junk := queries received for non-authoritative domains
- Distribution varies widely per zone
- ccTLDs: clouds send junk as all ASes do
- reduction in junk in junk levels to b.root-servers.net in 2020:
  - Proportionally, less junk from clouds
  - NSEC aggressive caching?
  - Chromium deployments now dominates root junk
Measuring Cloud Technology Adoption

- DNSSEC
- IPv4 vs IPv6
- UDP vs TCP

source: https://www.flickr.com/photos/anguskirk/4817305157
**DNSSEC**

- DNSSEC provides authenticity and integrity [1, 3, 2].
- Do clouds use it equally?
  - They need DS and DNSKEY records

**Queries Ratio**

- **A**: 0.02M / 1.1B
- **AAAA**: 11M / 460M

**w2020: .nl**

- Adoption measured by DNSKEY queries:
  - Microsoft: 0.02M / 1.1B
  - Cloudflare: 11M / 460M
IPv4 vs IPv6 Adoption

- Roughly 50/50%: Google, Cloudflare
- More IPv6: Facebook (2019 onwards)
- Very little IPv6: Microsoft, Amazon

<table>
<thead>
<tr>
<th>Year</th>
<th>.nl IPv4</th>
<th>.nl IPv6</th>
<th>.nz IPv4</th>
<th>.nz IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>0.66</td>
<td>0.34</td>
<td>0.61</td>
<td>0.39</td>
</tr>
<tr>
<td>2019</td>
<td>0.49</td>
<td>0.51</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td>2020</td>
<td>0.52</td>
<td>0.48</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td>Amazon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2019</td>
<td>0.98</td>
<td>0.02</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>2020</td>
<td>0.97</td>
<td>0.03</td>
<td>0.96</td>
<td>0.04</td>
</tr>
<tr>
<td>Microsoft</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2019</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2020</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Facebook</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>0.52</td>
<td>0.48</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td>2019</td>
<td>0.24</td>
<td>0.76</td>
<td>0.19</td>
<td>0.81</td>
</tr>
<tr>
<td>2020</td>
<td>0.24</td>
<td>0.76</td>
<td>0.17</td>
<td>0.83</td>
</tr>
<tr>
<td>Cloudflare</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>0.54</td>
<td>0.46</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td>2019</td>
<td>0.57</td>
<td>0.43</td>
<td>0.56</td>
<td>0.44</td>
</tr>
<tr>
<td>2020</td>
<td>0.51</td>
<td>0.49</td>
<td>0.49</td>
<td>0.51</td>
</tr>
</tbody>
</table>
UDP vs TCP

- UDP dominates
- TCP for large queries
- Facebook does more TCP (from 2019 onwards).

Why?

<table>
<thead>
<tr>
<th></th>
<th>Year</th>
<th>.nl</th>
<th>.nz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UDP</td>
<td>TCP</td>
<td>UDP</td>
</tr>
<tr>
<td>Google</td>
<td>2018</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Amazon</td>
<td>2018</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>Microsoft</td>
<td>2018</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Facebook</td>
<td>2018</td>
<td>0.79</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>0.85</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>0.86</td>
<td>0.14</td>
</tr>
<tr>
<td>Cloudflare</td>
<td>2018</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>0.99</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>0.98</td>
<td>0.02</td>
</tr>
</tbody>
</table>

UDP and TCP queries proportion
Why Facebook queries more TCP than others

- 1/3 of Facebook queries: EDNS(0) UDP size < 1024
- Sometimes caused truncated answers
- TCP required afterward

CDF of EDNS(0) UDP message size for .nl (w2020).
Conclusion: Clouds ain’t all the same

- DNS concentration:
  5 Clouds, 1/3 of ccTLD queries
- Technology adoption varies significantly
  - DNSSEC
  - Transport
  - Routing
- Centralization
  - Pro: new security feature deployments
    benefits many users all at once
  - Con: if it breaks, it can
    affect many users all at once

Questions?

real-world cloud types

Paper (IMC2020):
Download it here
DNS Security Introduction and Requirements. 

Protocol Modifications for the DNS Security Extensions. 

Resource Records for the DNS Security Extensions.

[4] BORTZMEYER, S.

DNS Query Name Minimisation to Improve Privacy.
[5] HARDAKER, W.

What’s in a name?
https://blog.apnic.net/2020/04/13/whats-in-a-name/.

[6] THOMAS, M.

Chromium’s impact on root dns traffic.