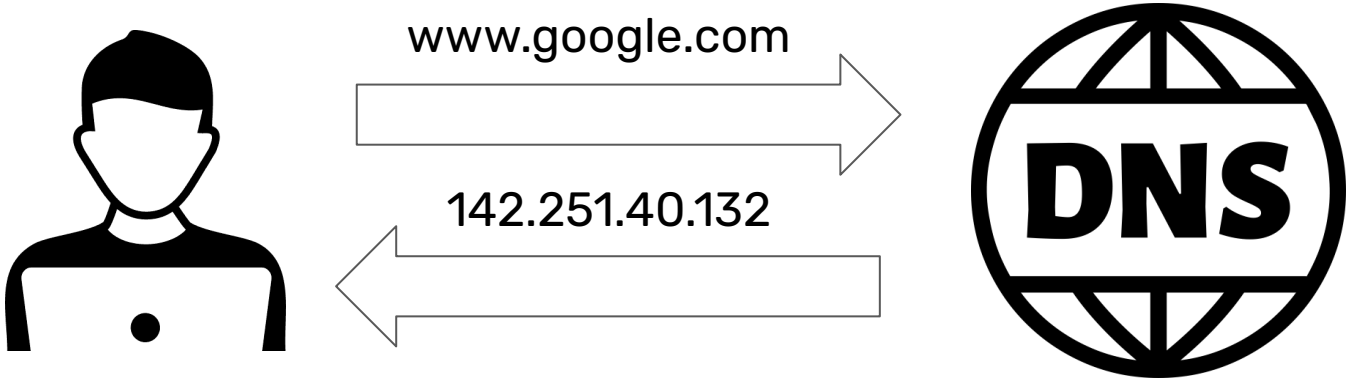
Understanding DNS Query Composition at B-Root

Jacob Ginesin, *Northeastern University,* Jelena Mirkovic, *USC/ISI*

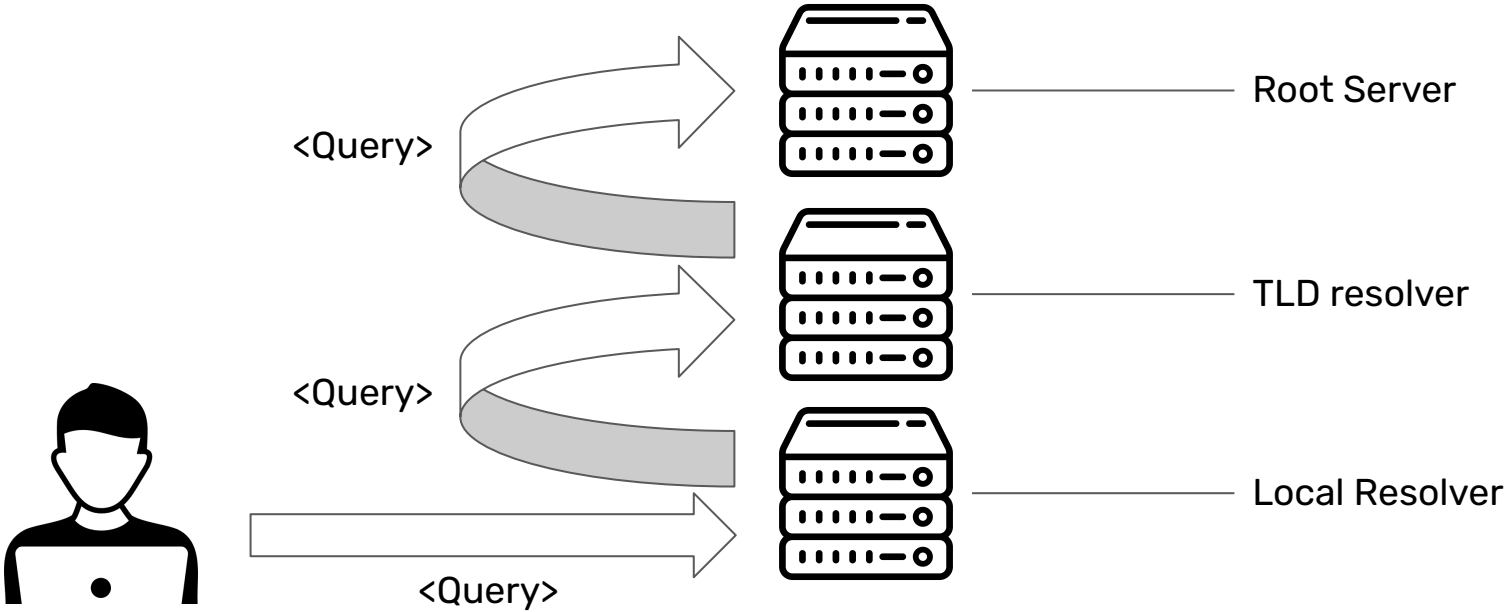
Background - What is DNS?

- The Domain Name System (DNS) is the internet's system for mapping an alphanumeric web address (e.g., www.google.com) to their respective IP address (e.g., 12.34.56.78)



Background - How Does DNS work?

- The DNS is a **Distributed Database**
  - DNS consists of thousands of servers, managed by hundreds of organizations
- DNS is **Hierarchical**



Background - Why Study The DNS?

- DNS is **critical internet infrastructure** - without it, the internet wouldn’t function
- Because of how important DNS is to the internet, it’s vital we understand its behaviors and specifics
- Little previous work in the space

Previous Work - A Day at the Root of the Internet

- “A Day at the Root of the Internet” - data analysis from 2007-2008 spanning 8 root servers
  - Did not include B-Root
- Found less than 2% of queries to root servers are by their definition, valid

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

Our Approach - Data

- To attempt to enumerate the behaviors of the DNS, we analyzed historical behavior at B-Root - a DNS root server
- B-Root is one of thirteen DNS root servers
- We sampled and analyzed traces at B-Root collected between 2013-2022 through the “A Day in the Life of the Internet” (DITL) project

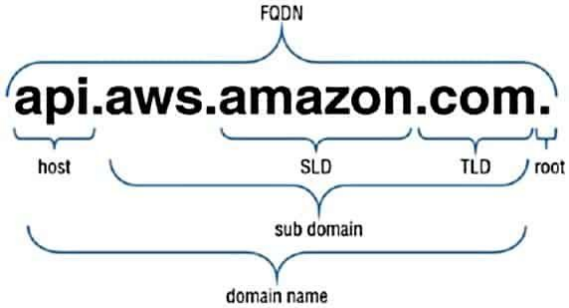
Information Sciences Institute @ USC (B-Root’s home!)

Our Approach - General Goals

- We want to quantify query types, top senders, and other interesting qualities of B-Root's data
- We want to identify historical trends
- We want to place all queries into disjoint, meaningful categories

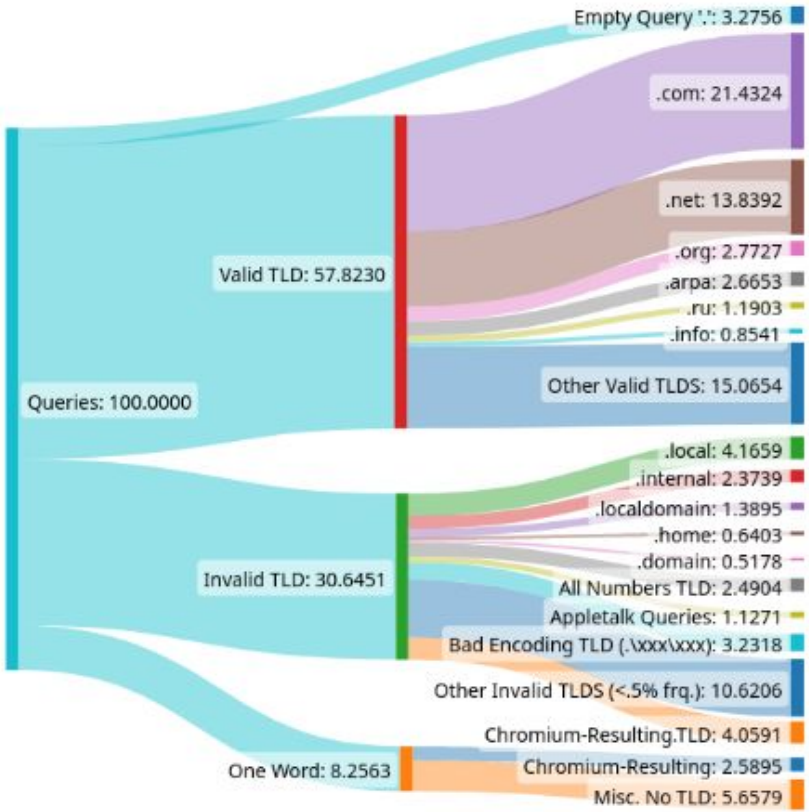
Our Approach - Classification Method

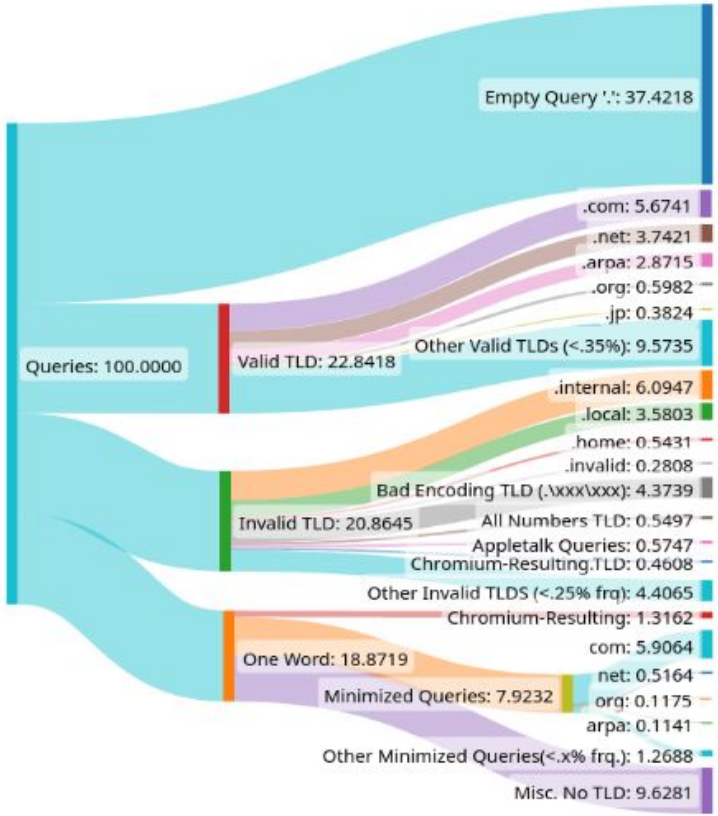
- To classify all queries into mutually exclusive and all-encompassing categories, we consider the domain name query structure (see: RFC1034)
- Moving from right to left on a query, there are only four possible options
  - The query is empty - “.”
  - The query has no TLD - “foo.”
  - The query has a valid TLD - “google.com.”
  - The query has an invalid TLD - “google.local”
- Within these four categories, we additionally classified types of queries we deemed interesting



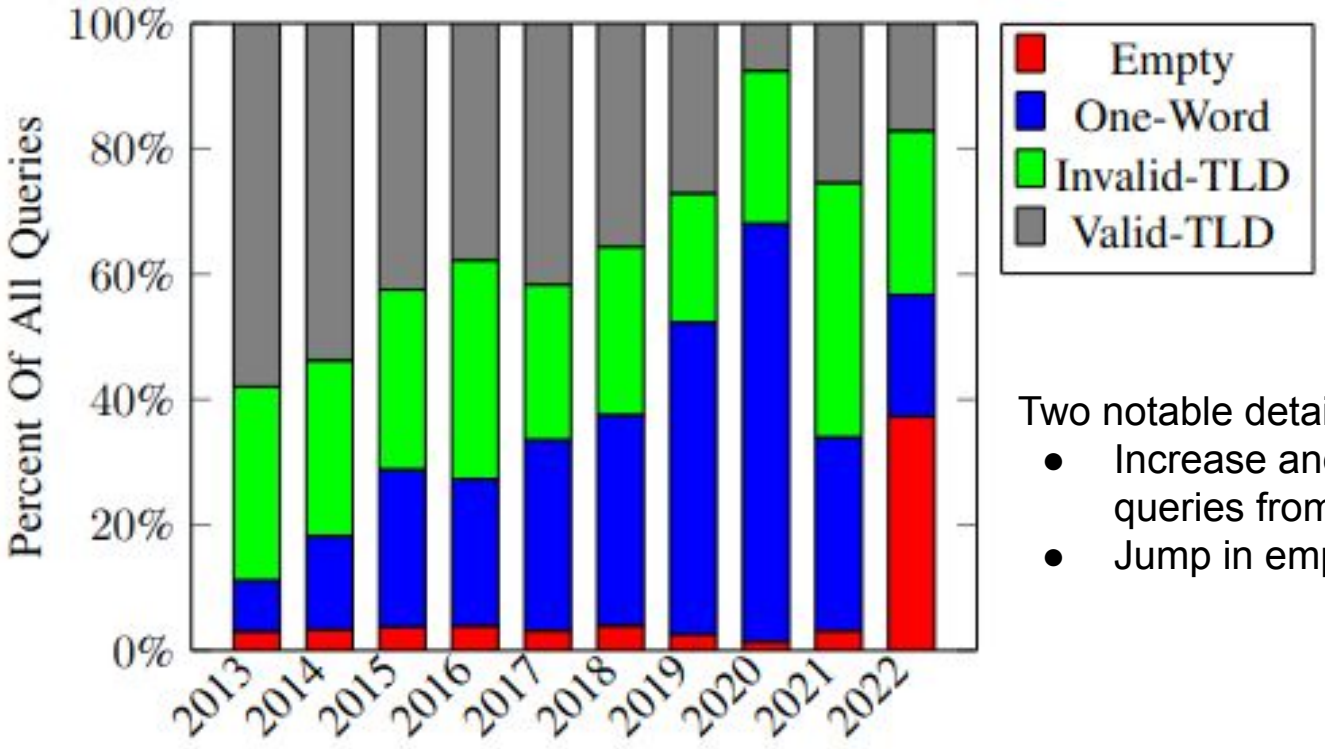
host.SLD.TLD.root

Results - General Query Classification, 2013 vs 2022





Results - General Classification from 2013 to 2022

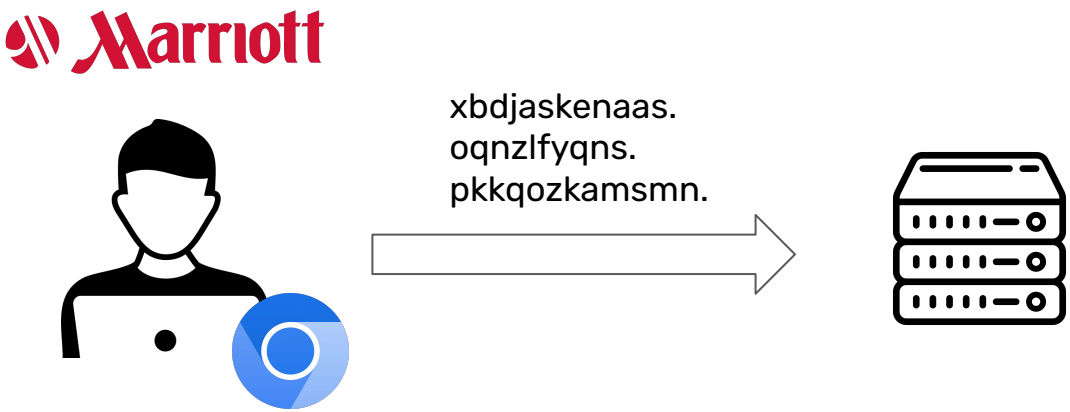


Two notable details:

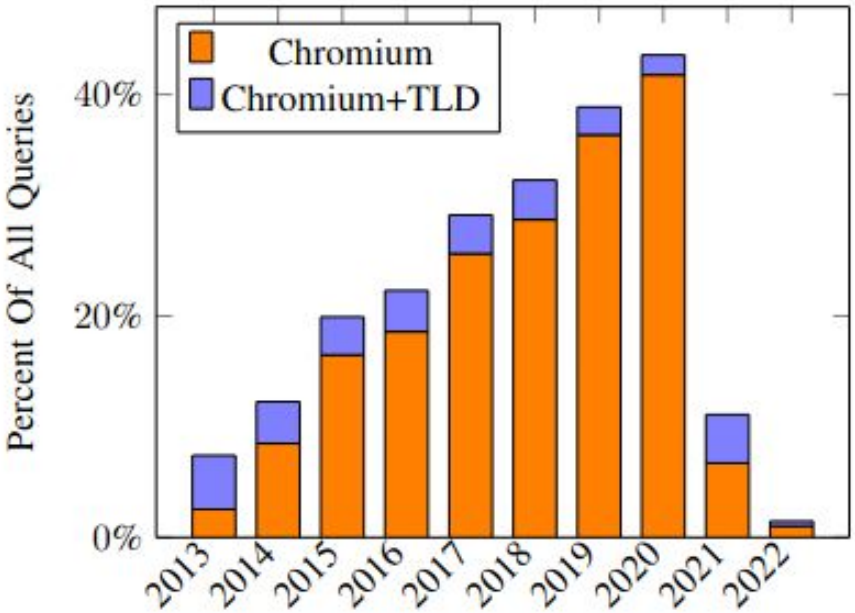
- Increase and decrease in one-word queries from 2013-2022
- Jump in empty queries in 2022

Results: Chromium-Originating Queries (Background)

- Chromium detects the presence of captive portals by sending three randomly generated queries to DNS servers
- This results in tons of garbage queries at B-Root
- Issue fixed in 2020 with the Chromium 87 release



Results - Chromium-Originating Queries



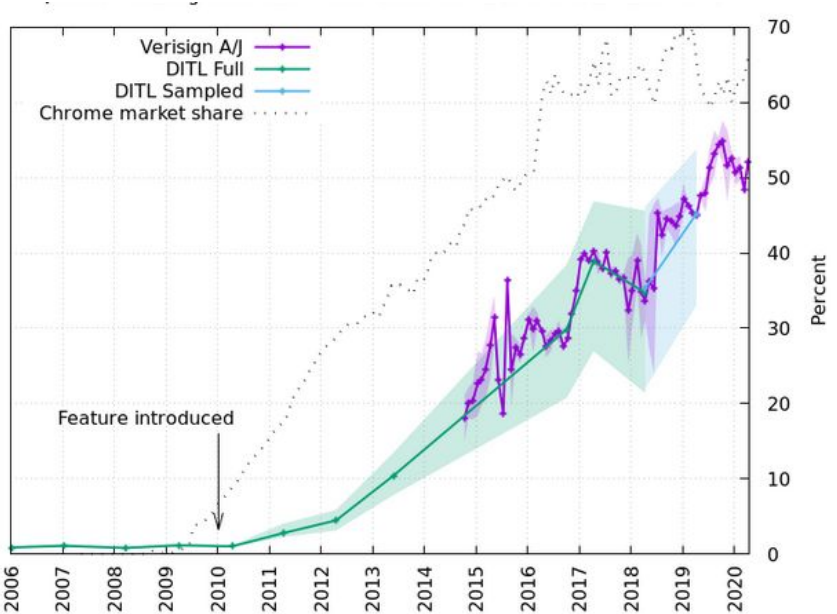
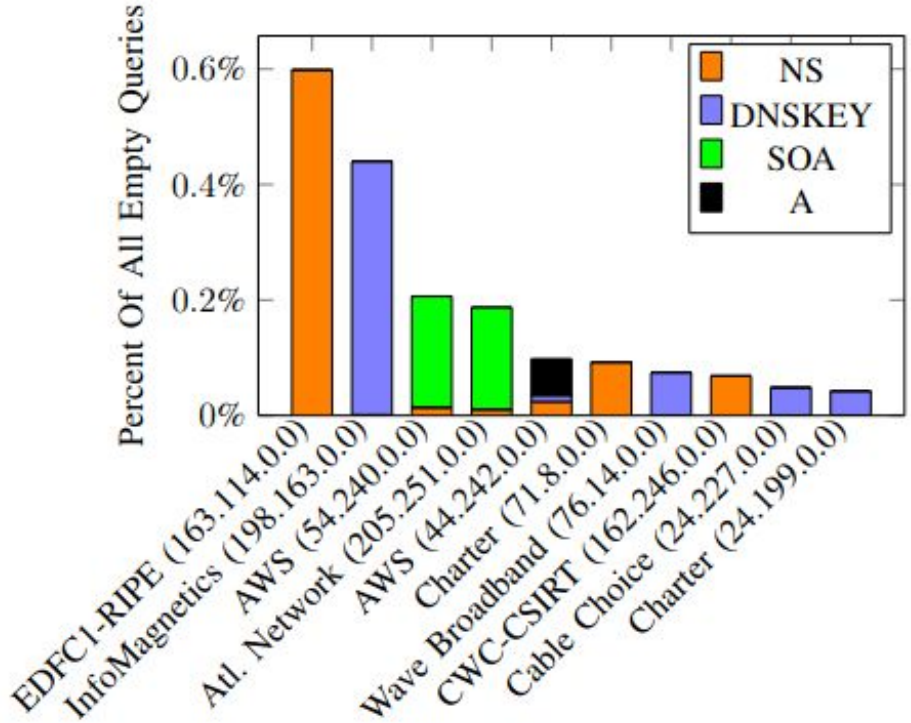


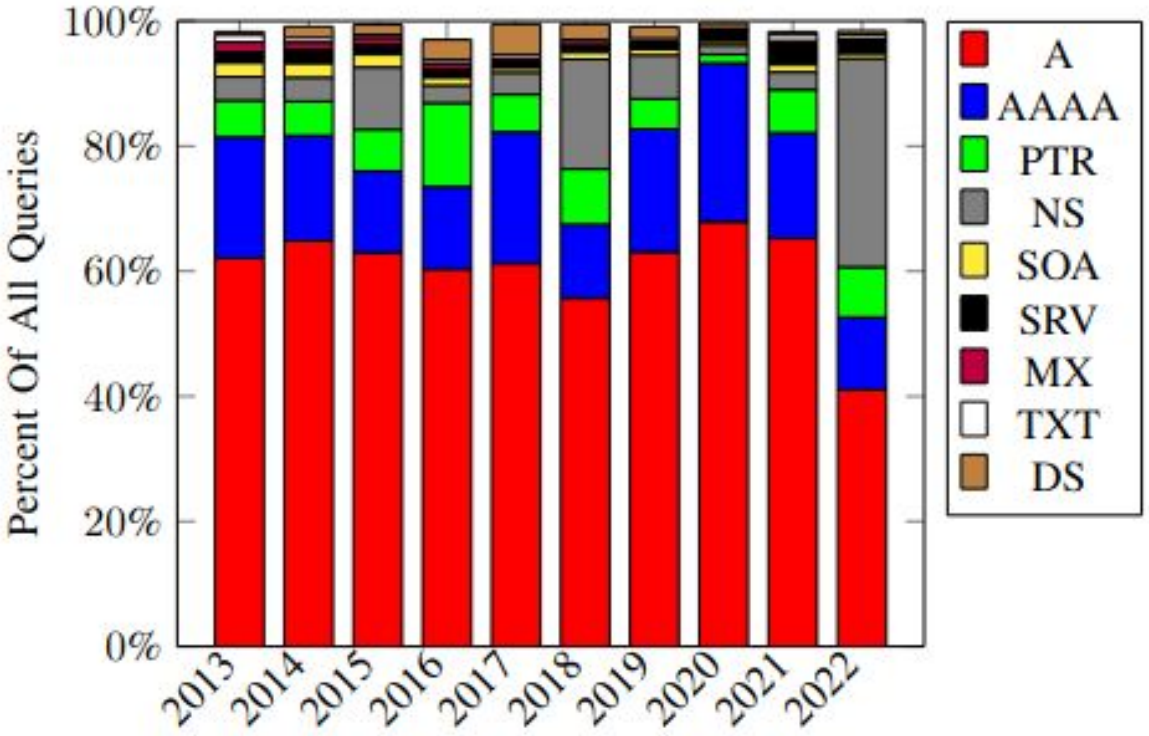
Image Source: blog.apnic.net

Results - Empty Queries

- Senders are decentralized
- 97% of empty queries in 2022 were of type NS (name server)
- Likely priming queries - a new DNS feature introduced in 2017 (see: RFC8109)
    - Priming queries are specified to have type “NS”



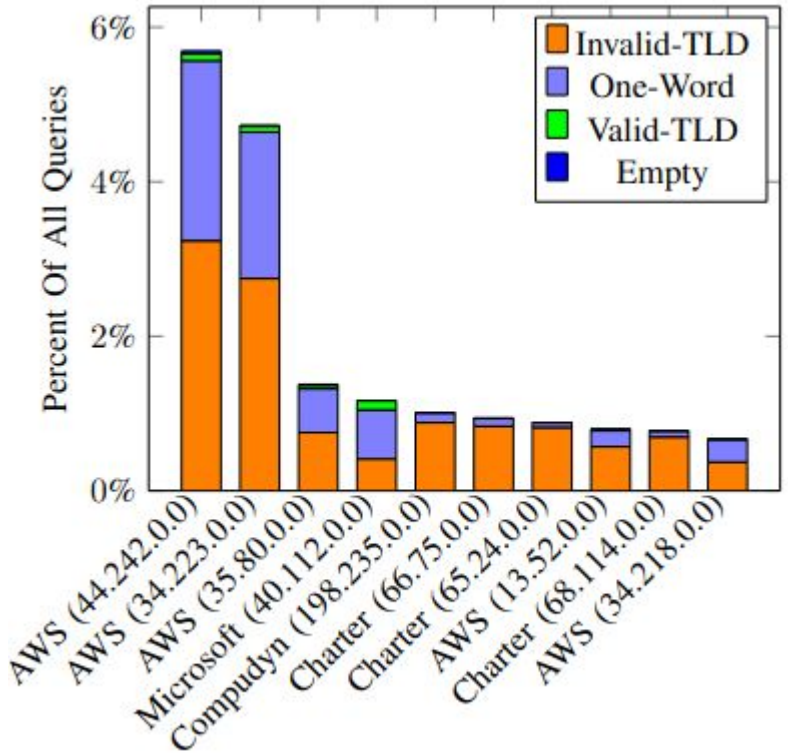
Results - Historical Query Types, 2013-2022



- A: IPV4
- AAAA (4A): IPV6
- PTR: Reverse Mapping Query
- NS: Name Server Query
- SOA: Start-of-authority
- SRV: host+port
- MX: Mail
- TXT: Text records
- DS: Stores public keys

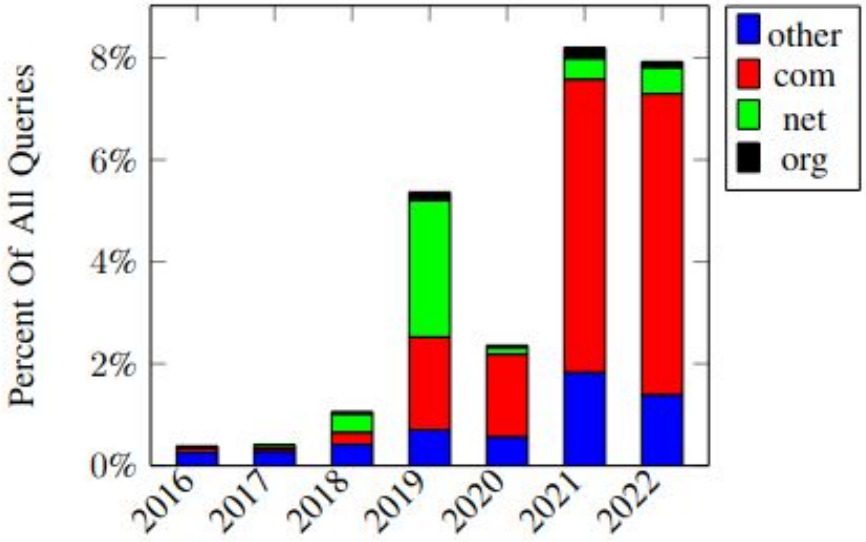
Results - Top Query Senders in 2022

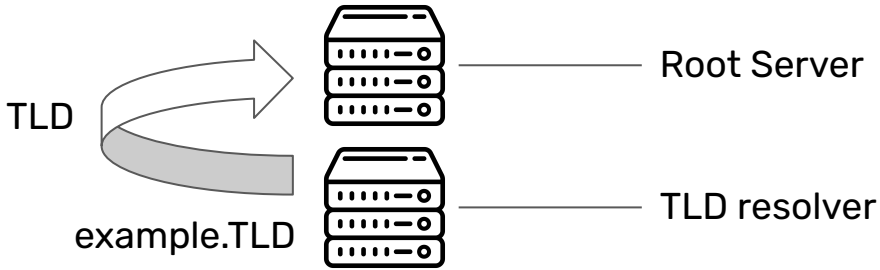
- Amazon consists of ~20% of query traffic to B-Root
- Most top query senders send invalid queries



Results - Query Minimization Queries

- Query Minimization was introduced in 2016 (RFC7816)
- Minimizes the amount of data sent as query is recursed
- These queries appear as one-word valid TLDs (e.g, “com.”)





Conclusion

- We investigated 10 years of B-Root's DNS traces and categorized longitudinal trends
- We categorized a high volume of priming queries in B-Root's 2022 data
- We only categorized traffic at B-Root - in the future, we would like to expand our analysis to all 13 root servers.

Fin!

</> full paper: https://jakegines.in/research/dns-b-root

 personal site: https://jakegines.in