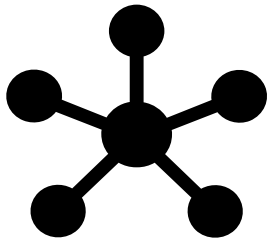
Trufflehunter: Cache Snooping Rare Domains at Large Public DNS Resolvers

Audrey Randall, Enze “Alex” Liu, Gautam Akiwate, Ramakrishna Padmanabhan, Geoffrey M. Voelker, Stefan Savage, Aaron Schulman

UCSDCSE
Computer Science and Engineering

Categories of harmful Internet behavior

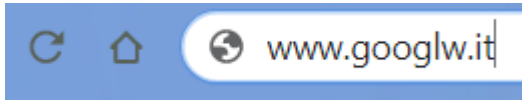
Spam Emails



Botnets

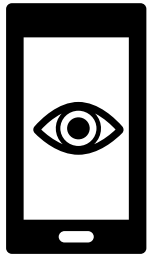
Malware

Common Internet abuse
(well studied)



Typo Squatting

Hack for Hire



Stalkerware

Rare Internet abuse
(sparsely studied)

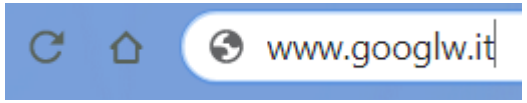
Categories of harmful Internet behavior

Spam Emails

Botnets

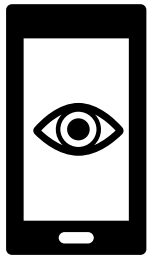
Malware

Common Internet abuse
(well studied)



Typo Squatting

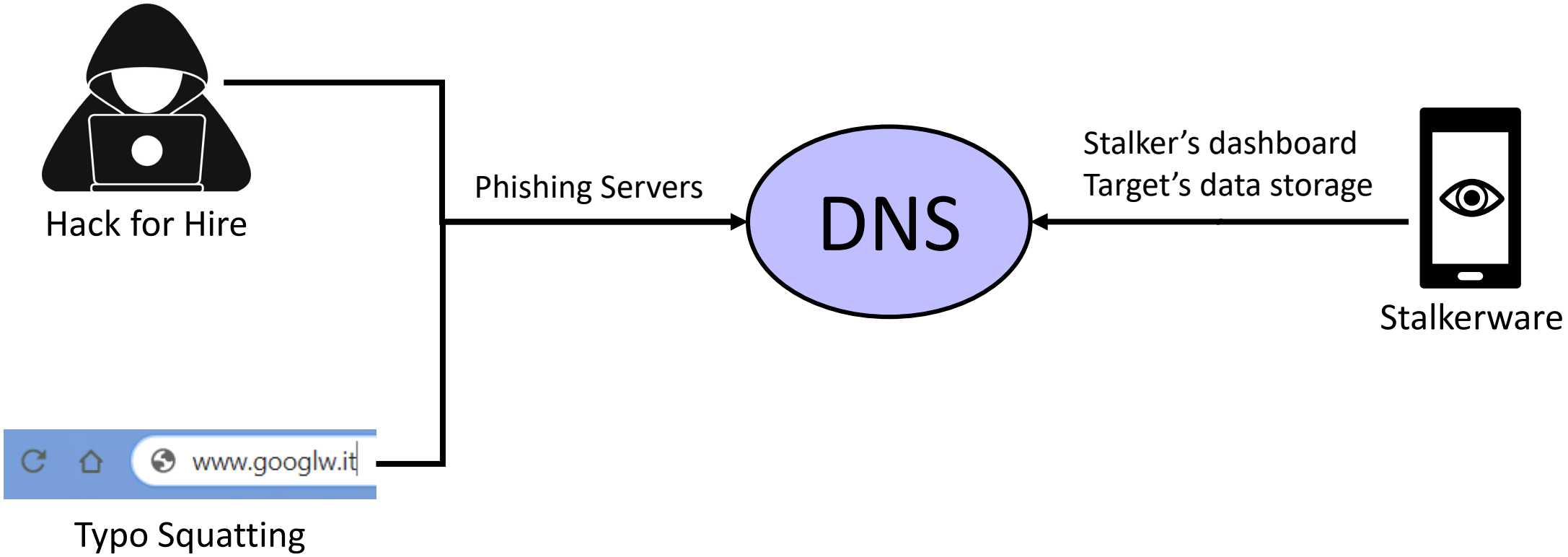
Hack for Hire



Stalkerware

Rare Internet abuse
(sparsely studied)

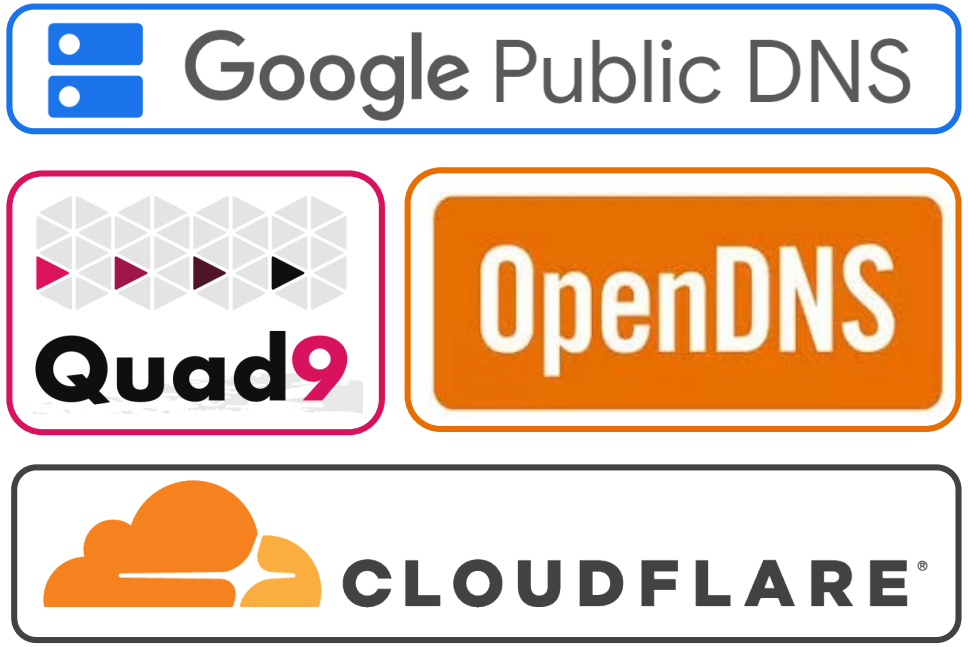
Categories of harmful Internet behavior



If you can observe the DNS, you can observe these behaviors.

New Era in DNS: Public Resolvers

- Public resolvers are gaining popularity
- Many users now use these resolvers by default
  - Google home routers go to 8.8.8.8
  - Cloudflare DNS is default on Firefox
  - NYC Public WiFi uses Quad9
- Can a third-party observer use these services to observe rare behavior?



Observing requests on public resolvers

Well-known technique: DNS cache snooping.

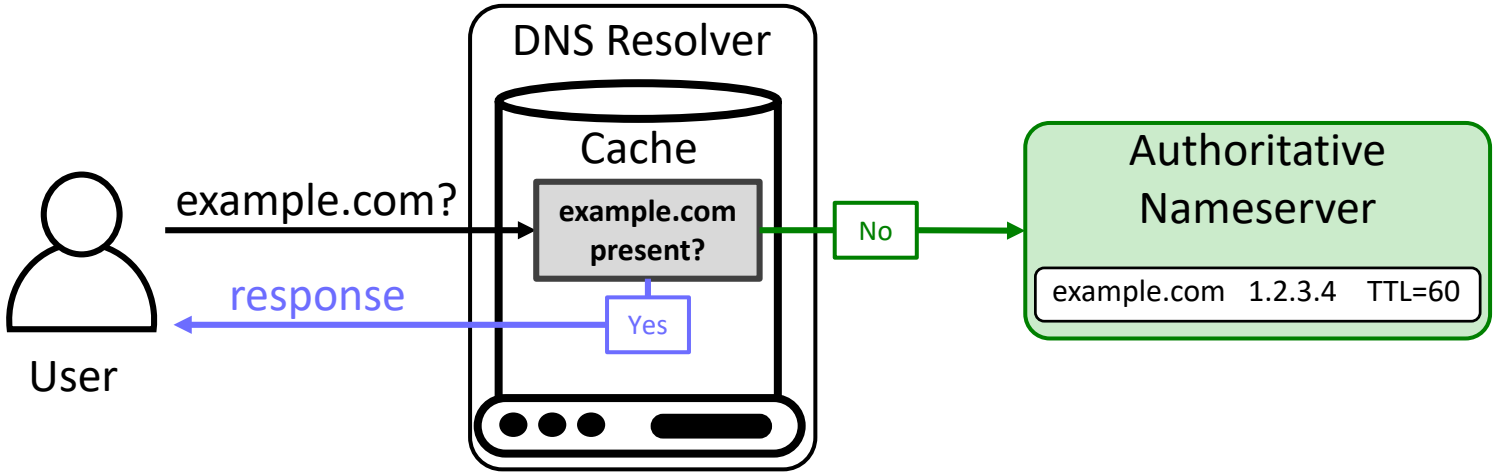
Previous work: Find resolvers by scanning Internet

- Open resolvers usually misconfigured home routers – privacy threat

On public DNS resolvers, it's privacy-preserving!

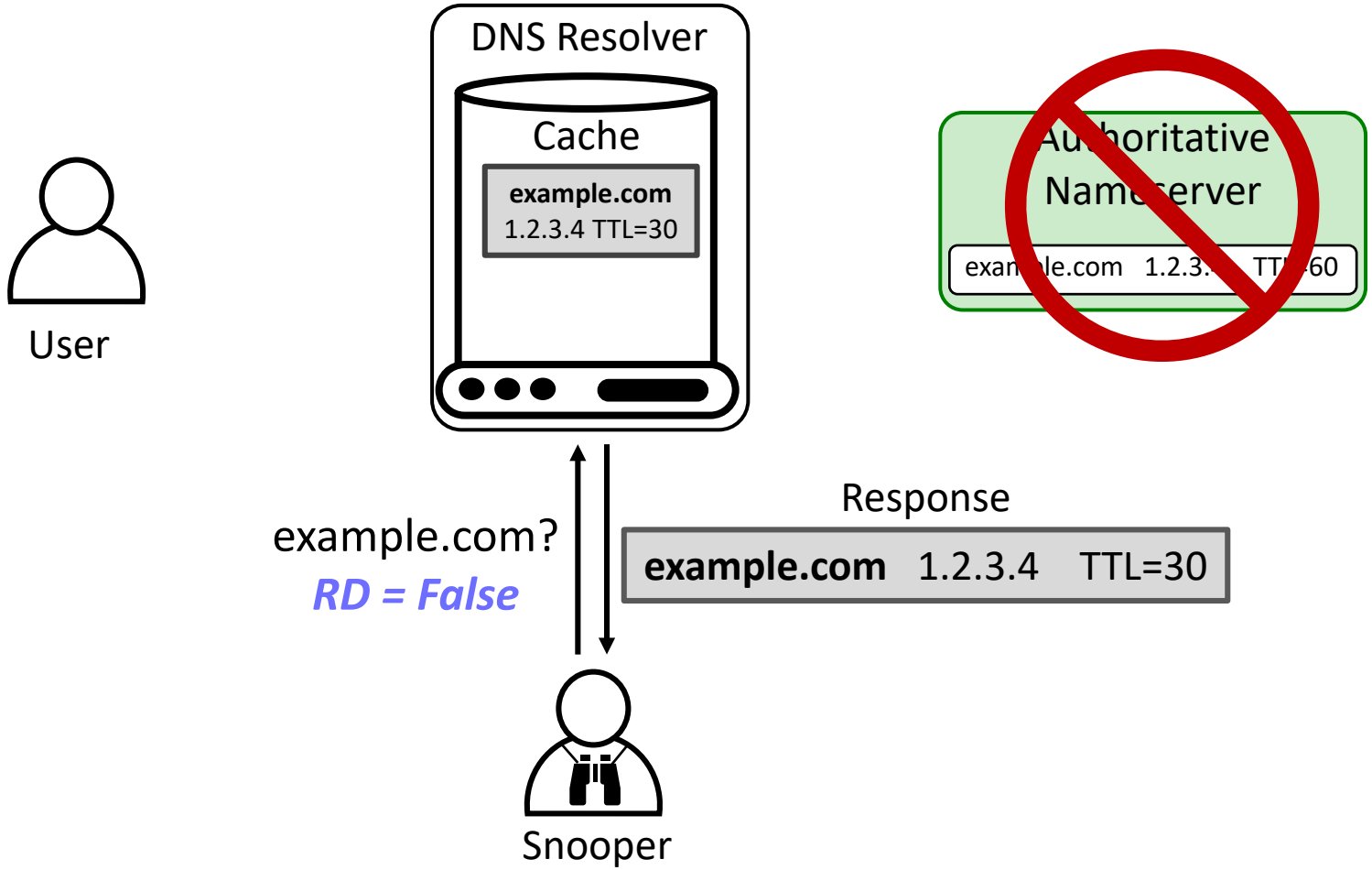
- Too many users to de-anonymize

Background: How Cache Snooping Works

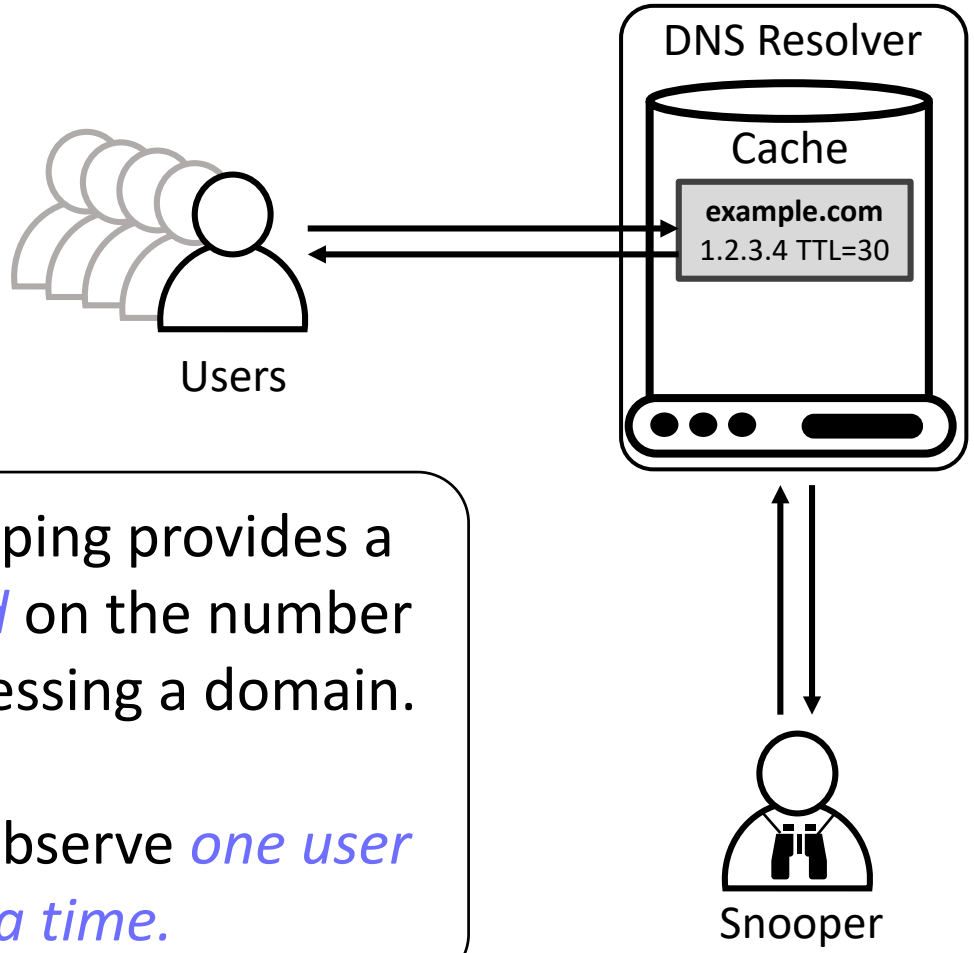


Snooper

Background: How Cache Snooping Works



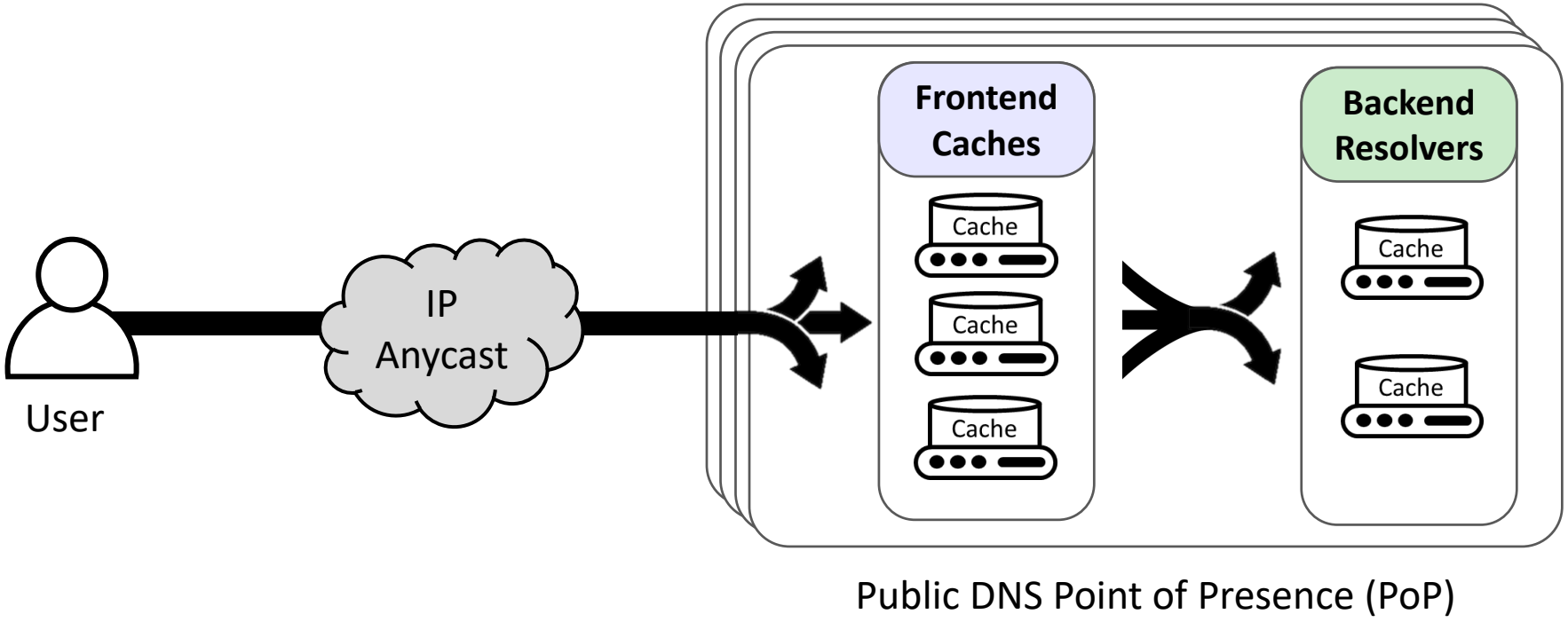
Background: How Cache Snooping Works



Cache snooping provides a *lower bound* on the number of users accessing a domain.

It can only observe *one user at a time.*

Simplified Public Resolver Cache Architecture

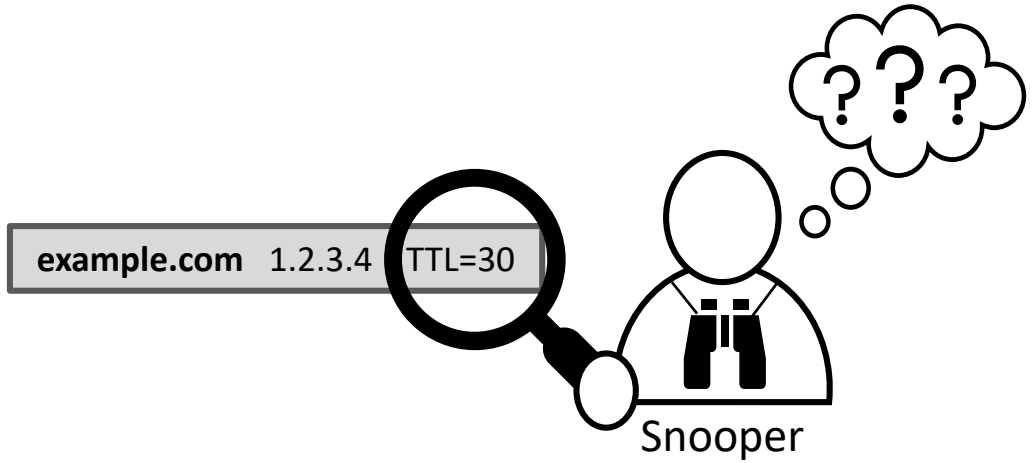


And it gets more complicated…

A snooper's ability to estimate depends on their understanding of the cache architecture.

- Snooper only sees TTL and timestamp of DNS query!

We reverse engineer a model of each resolver to make snooping possible.



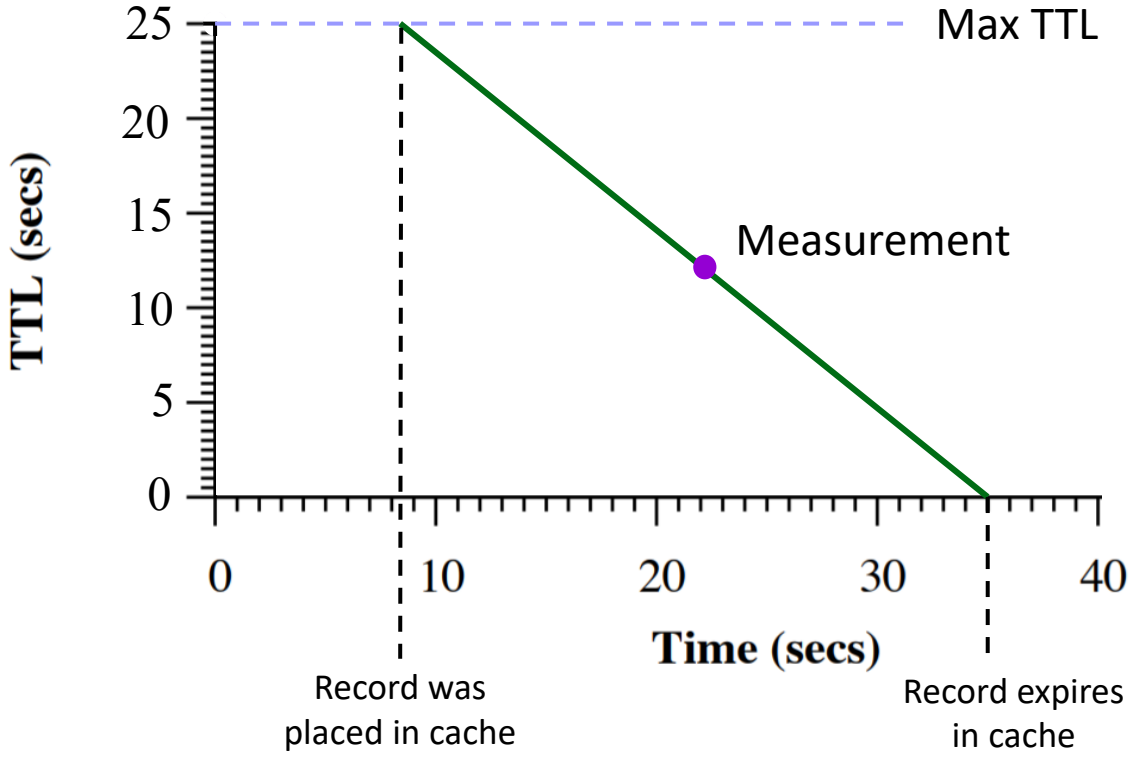
How We Modeled Cache Architectures

Experiment:

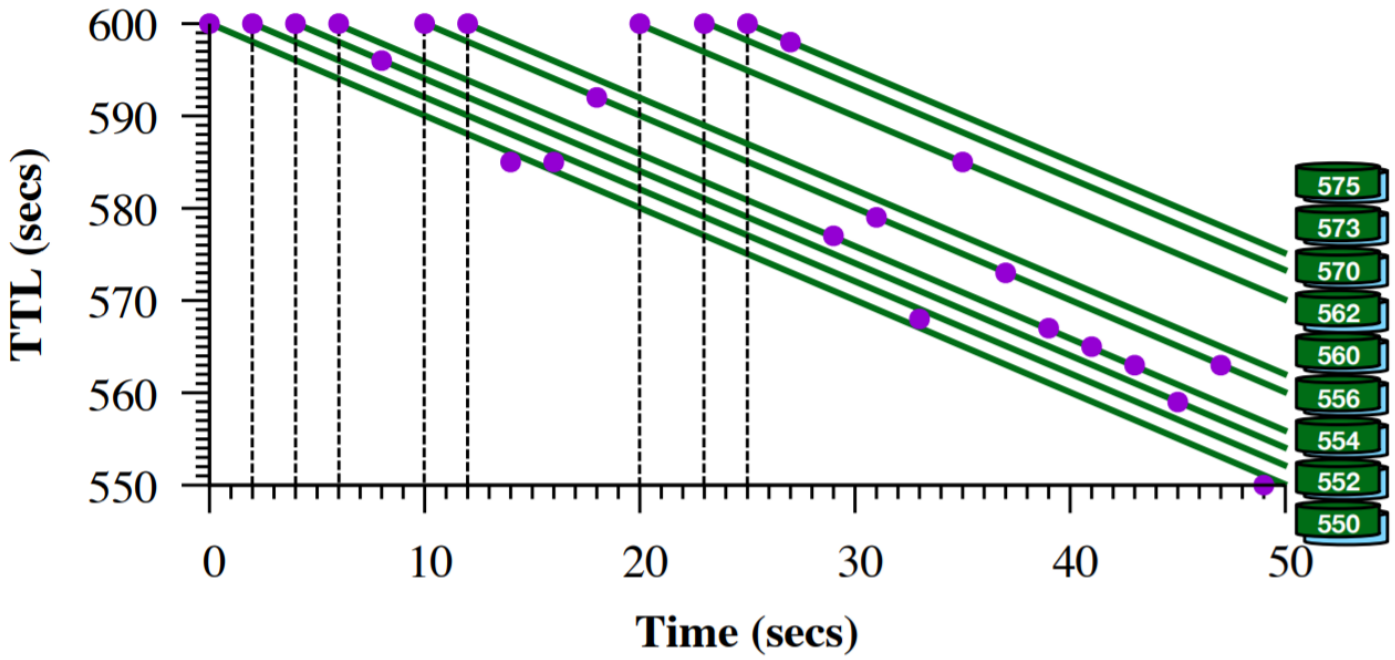
- Repeatedly query a resolver
  - Fill as many caches as possible
- Observe how our queries were cached: examine TTLs.
  - Some queries will hit the same cache, others will not

"TTL Line:" Model of how a TTL decreases in a cache.

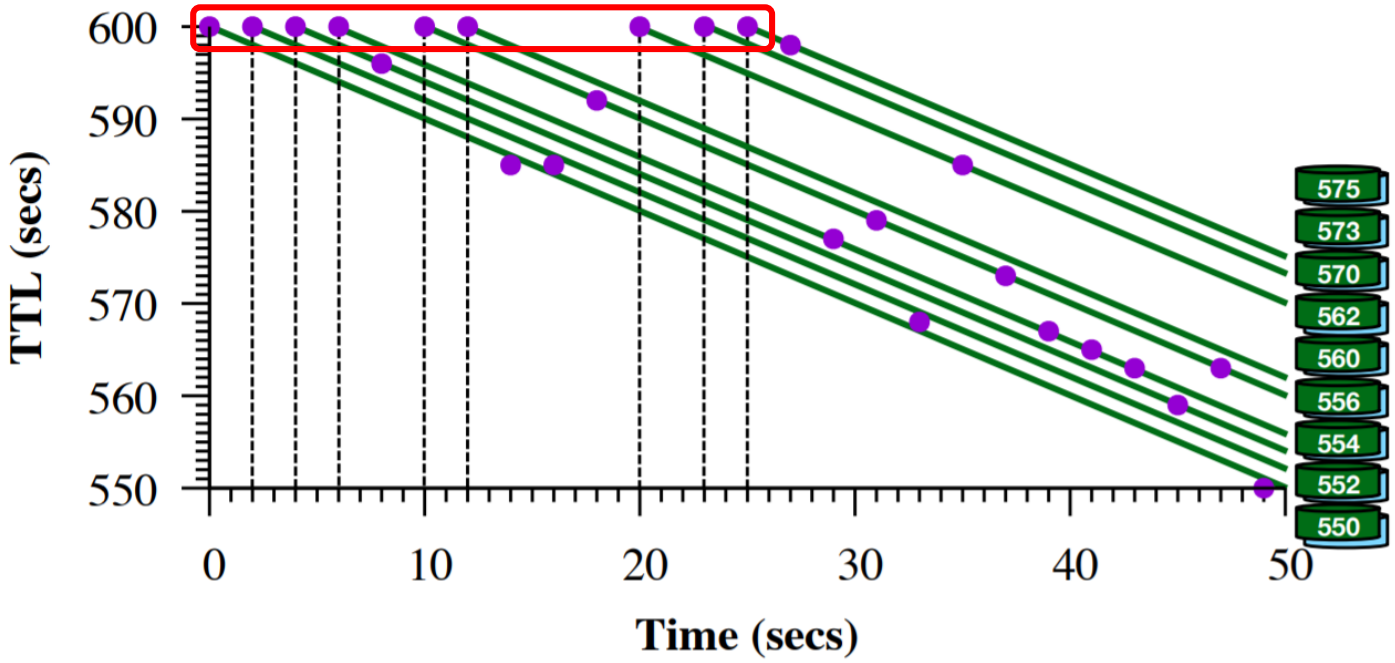
- Rate: one second per second.



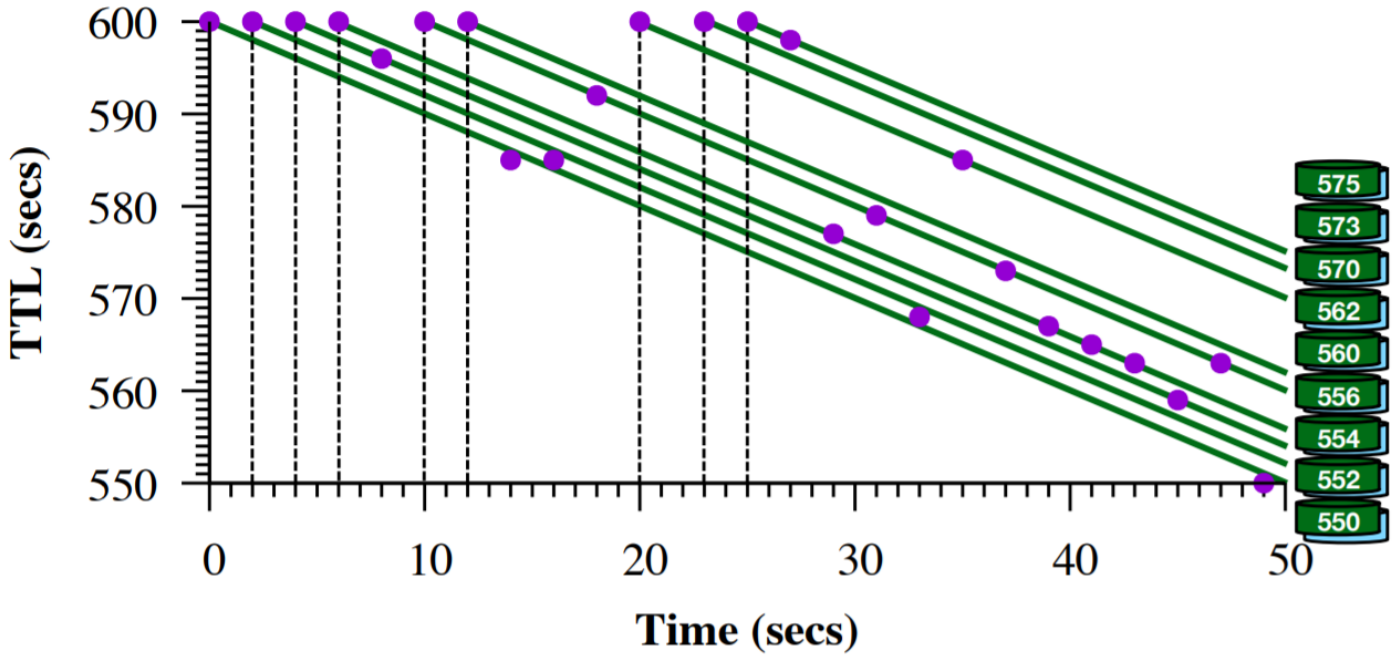
OpenDNS and Quad9

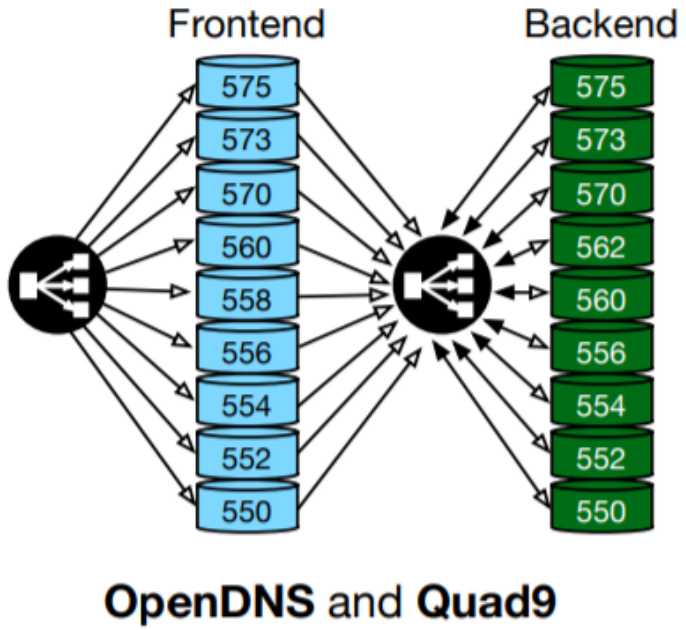


OpenDNS and Quad9



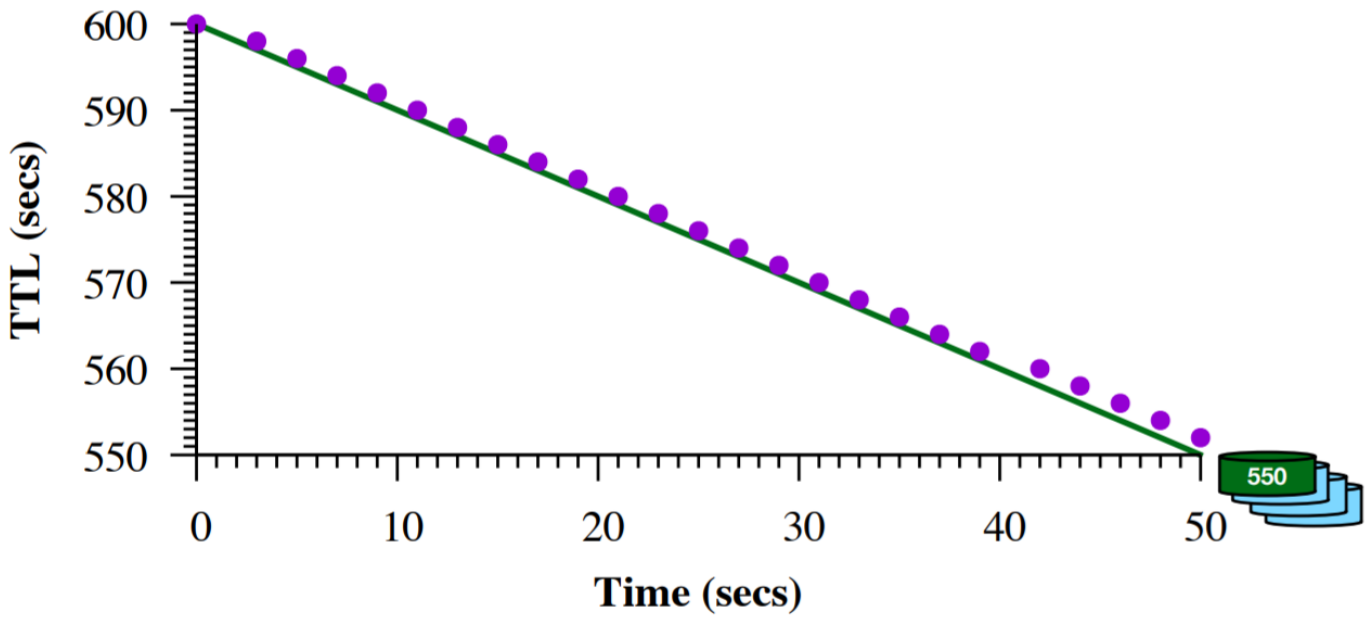
OpenDNS and Quad9



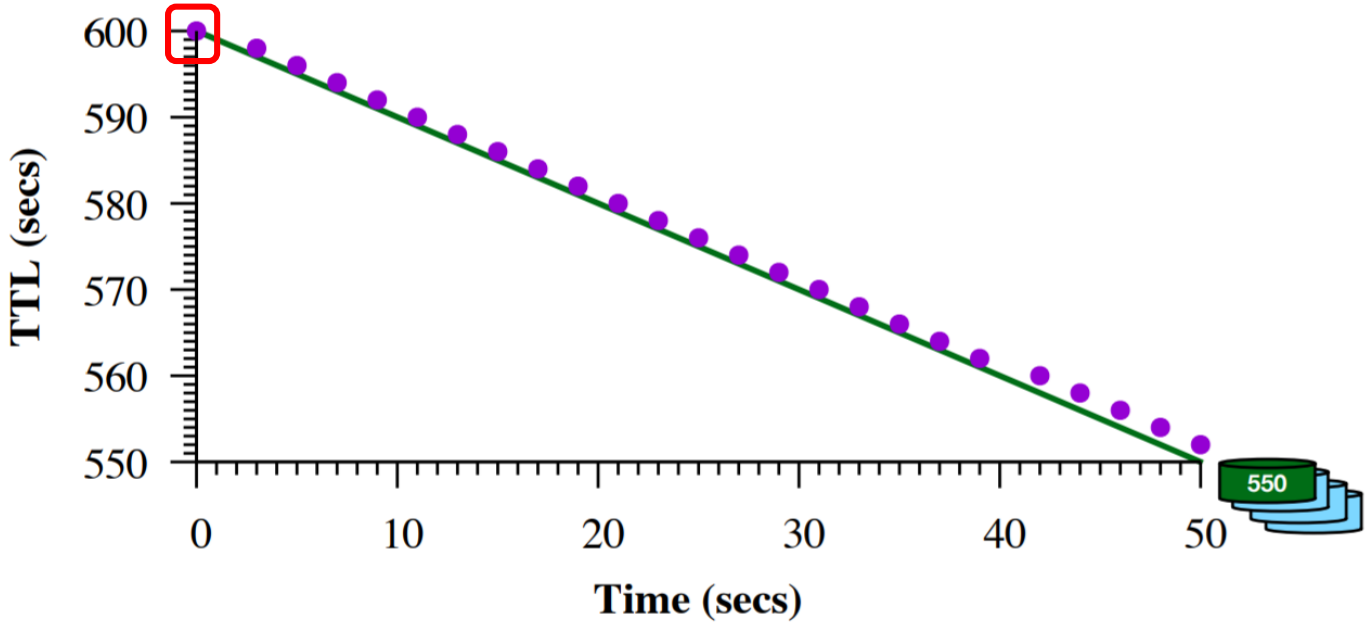


OpenDNS and **Quad9**

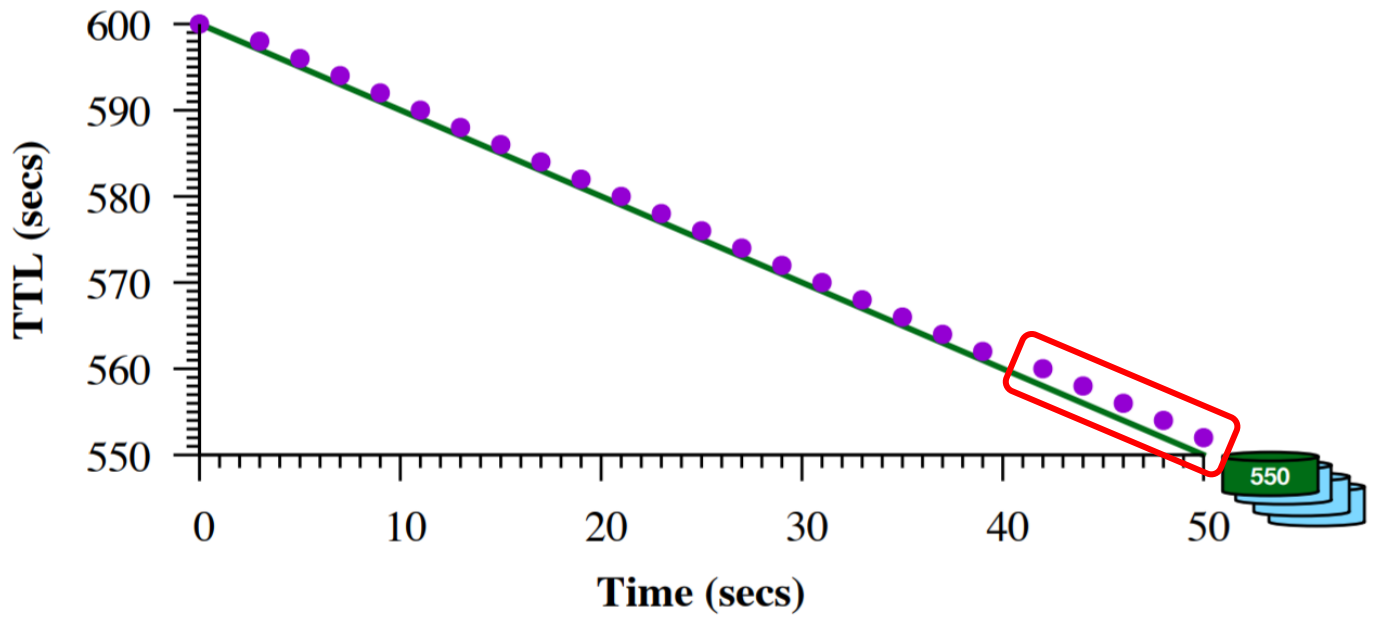
Cloudflare



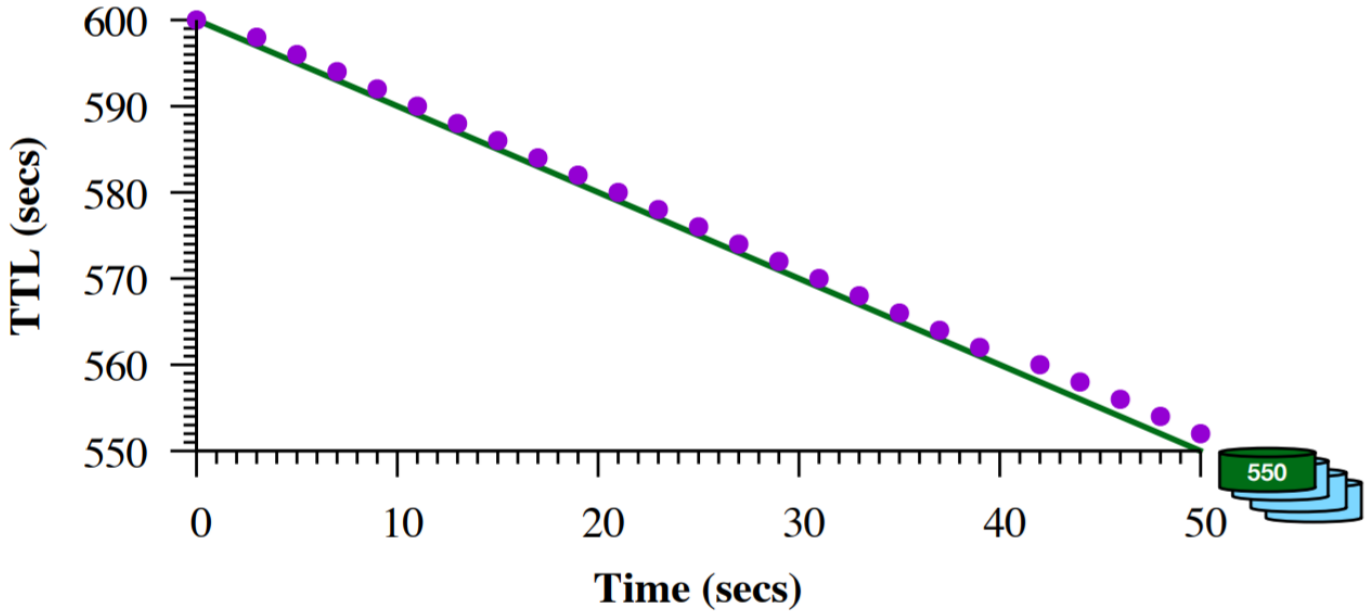
Cloudflare

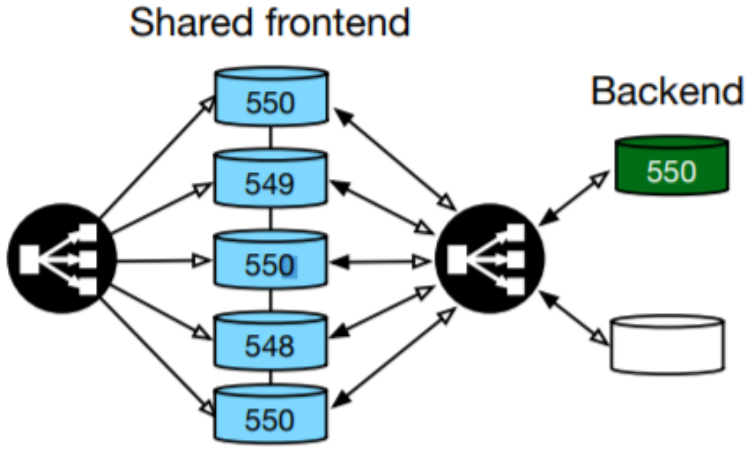


Cloudflare



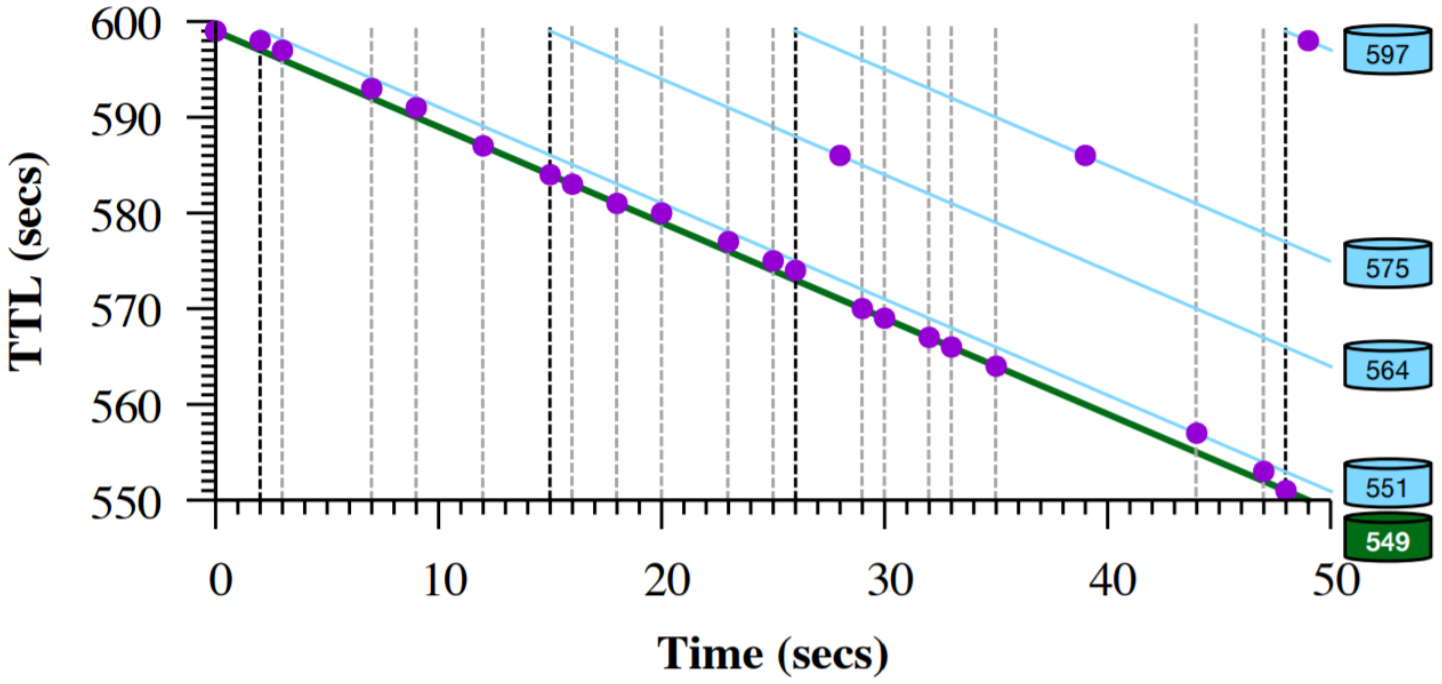
Cloudflare



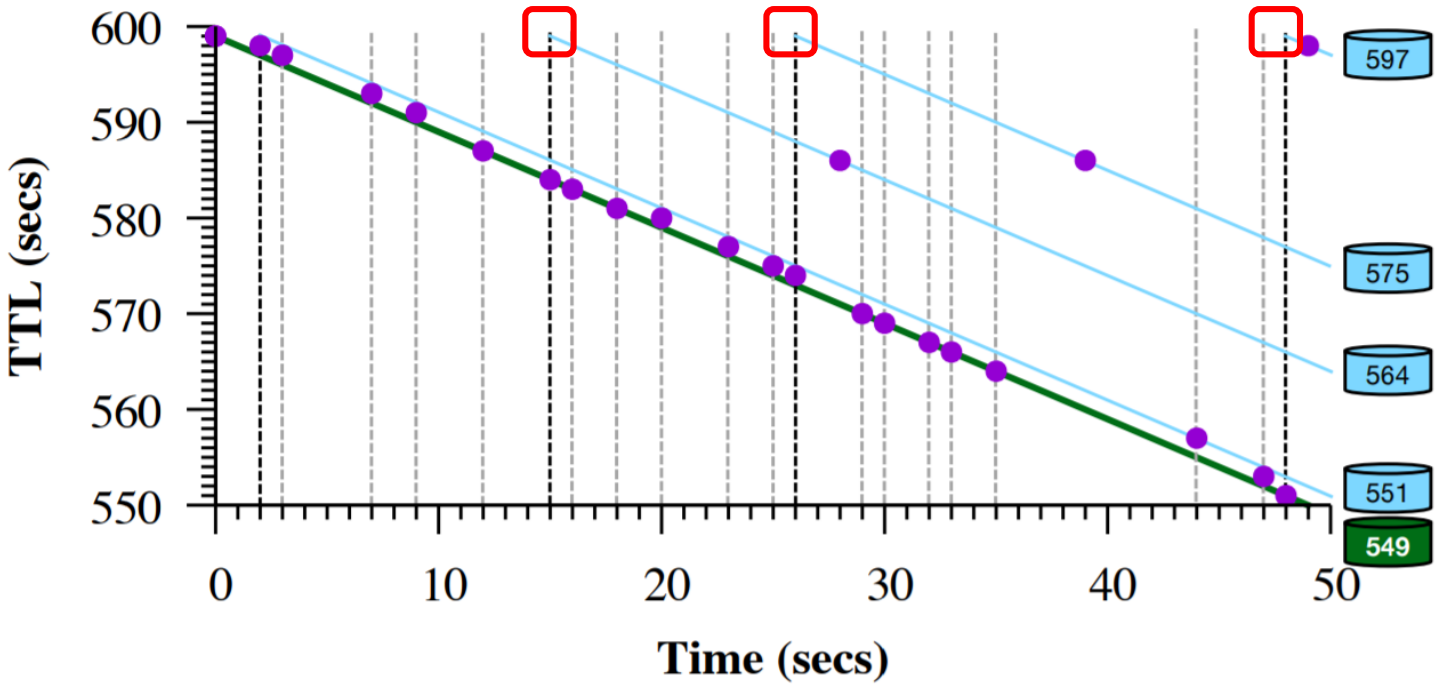


Cloudflare

And then there's Google Public DNS…

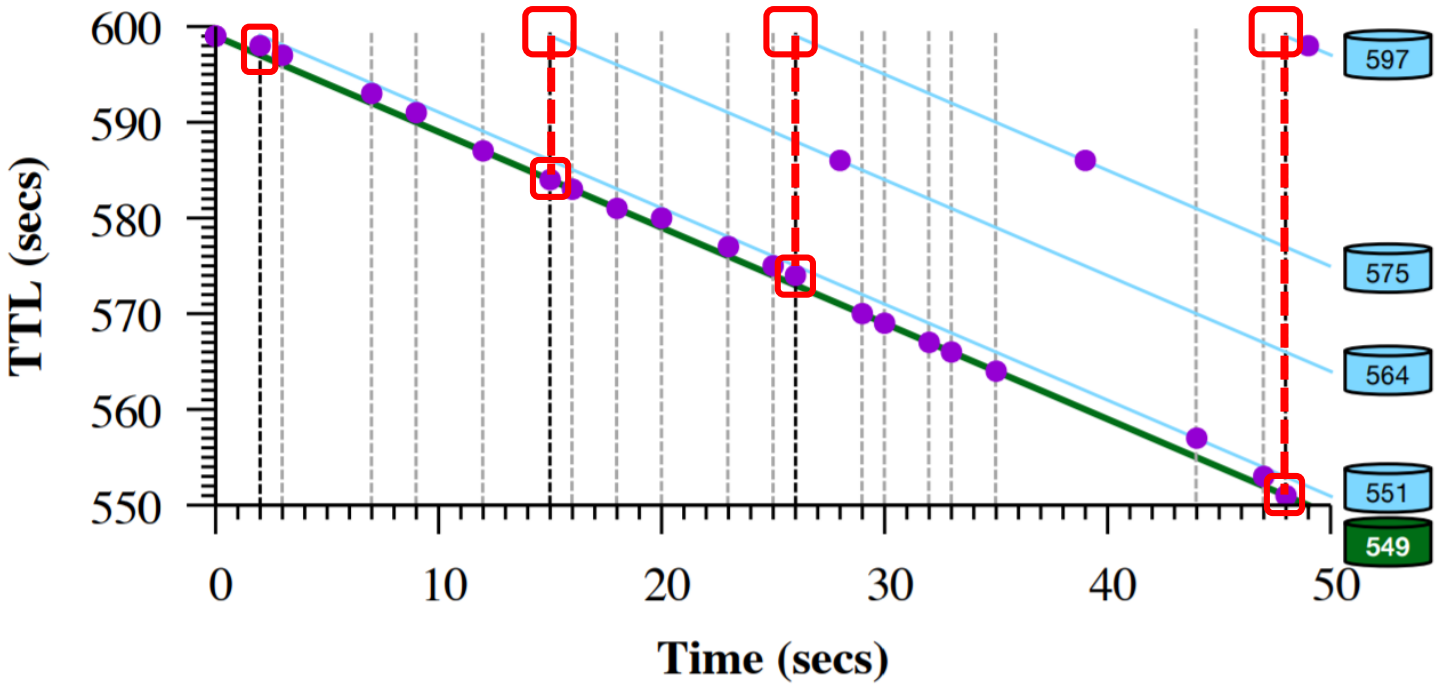


And then there's Google Public DNS…

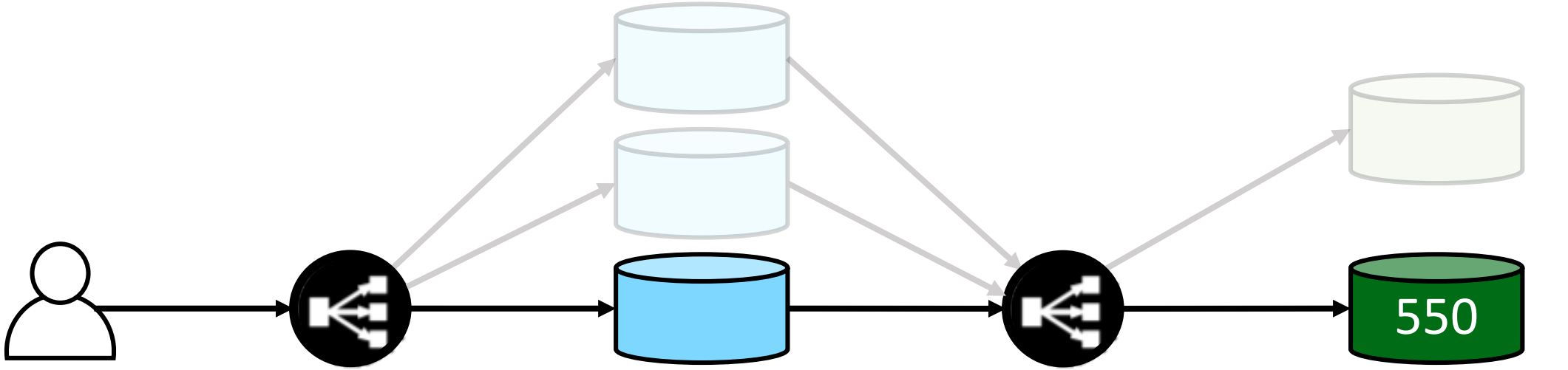


No measurements!

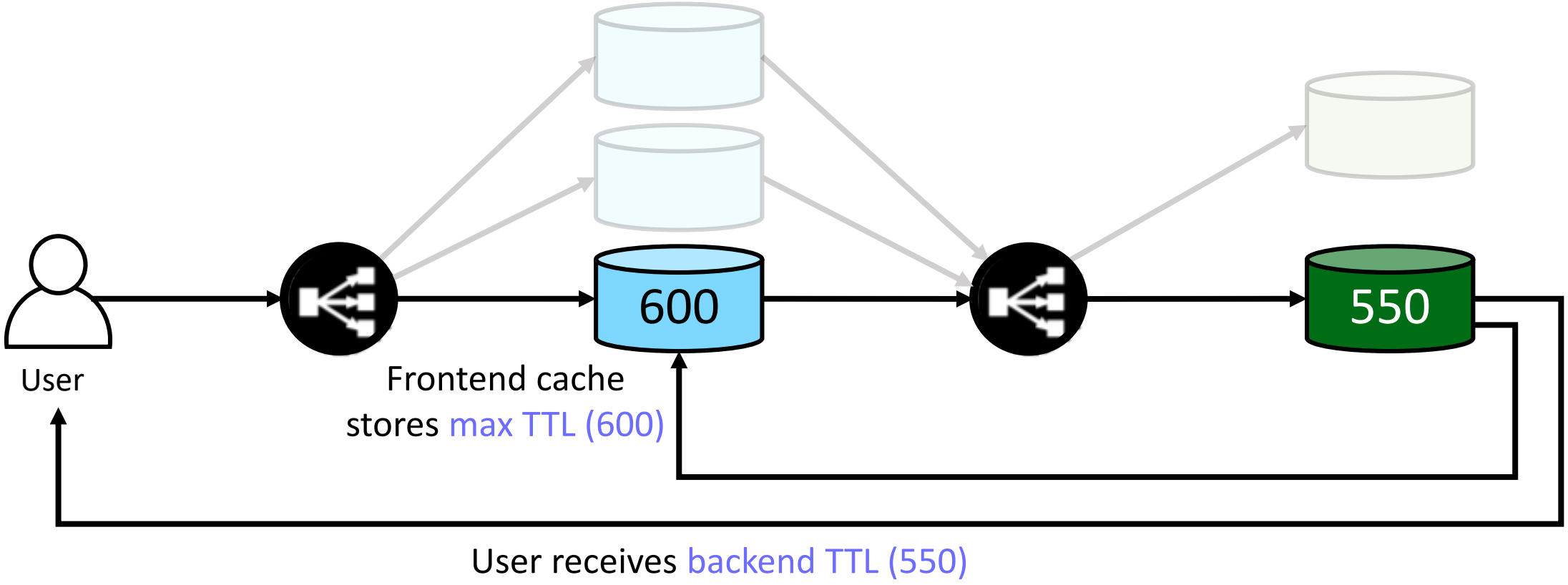
And then there's Google Public DNS…



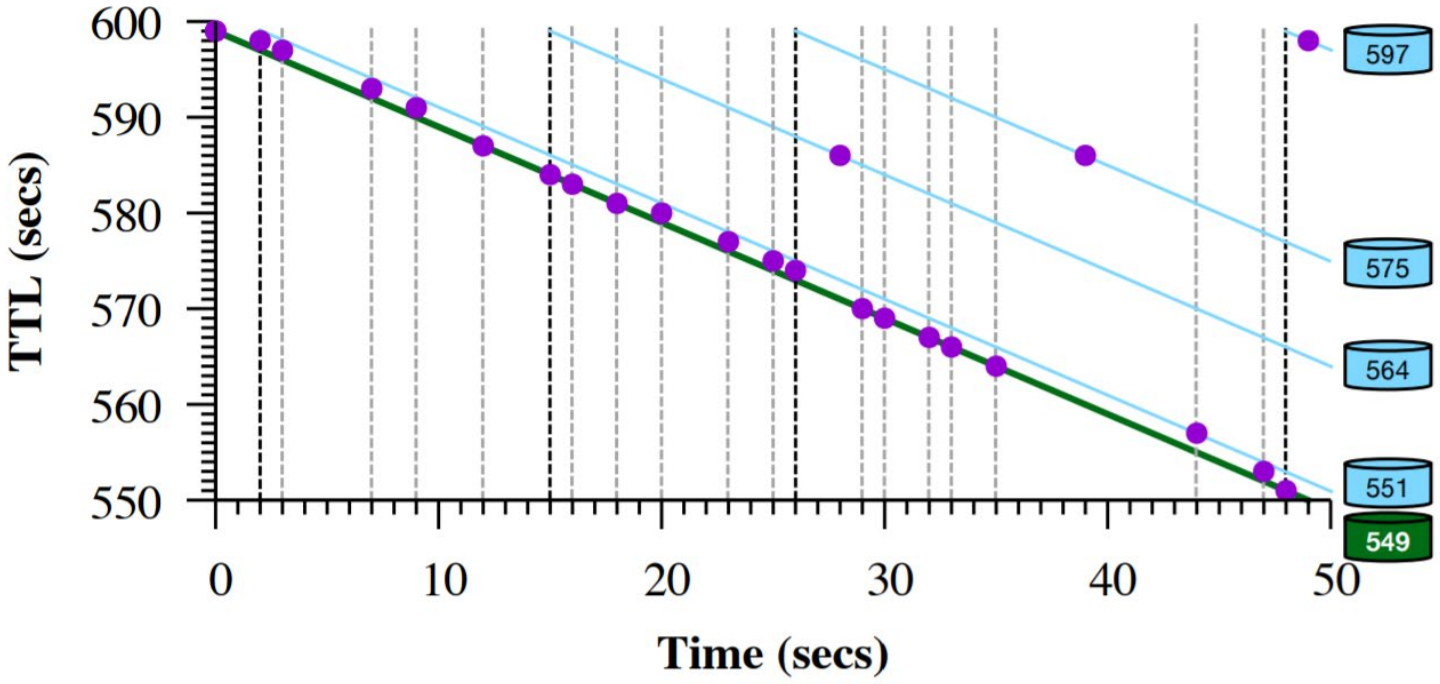
Google Public DNS: Dynamic Caching

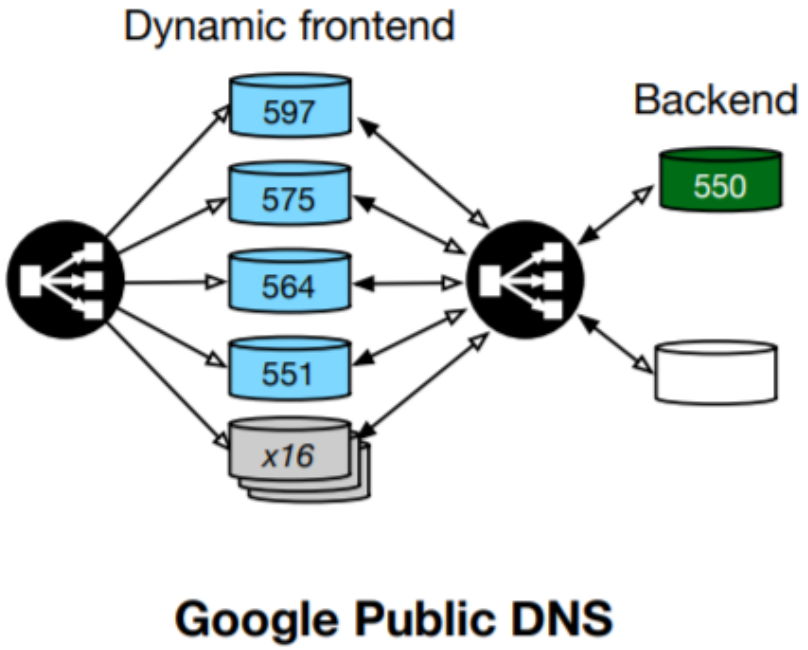


Google Public DNS: Dynamic Caching



Google Public DNS: Dynamic Caching





Trufflehunter

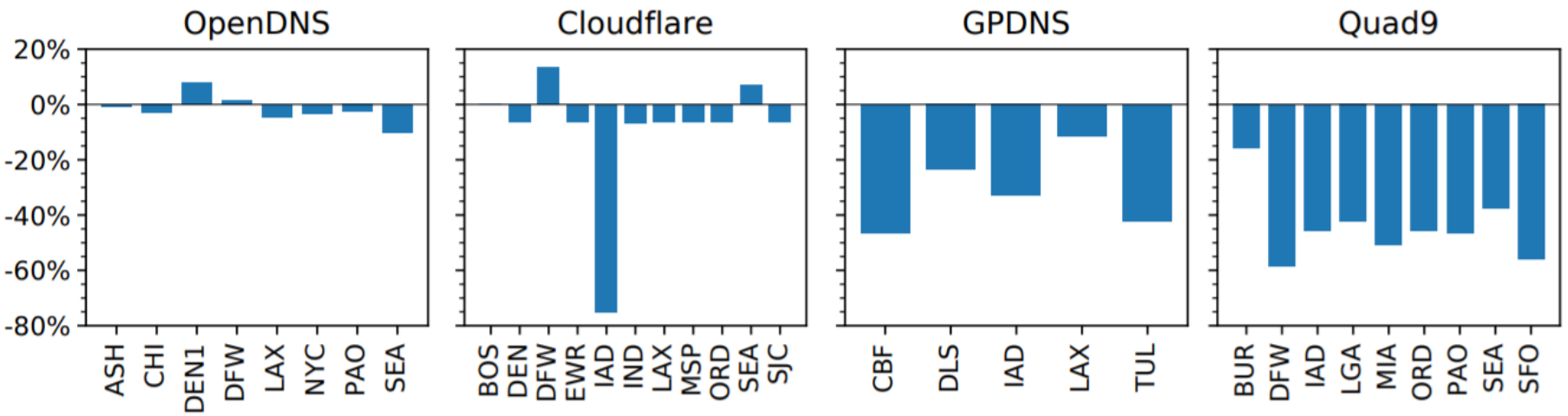
- Distributed measurement tool
  - Deployed on CAIDA's Ark project
- Sends DNS queries for domains of interest across the U.S.
- Interprets the responses according to our models to estimate counts of users
- Three months of data: March 6 – May 29 2020

How accurate is Trufflehunter at estimating filled caches?

Experiment:

- Place domain controlled by us into cache using ~900 RIPE Atlas probes
- Attempt to observe this domain with Trufflehunter
- Number of requests to our authoritative nameserver is true number of filled caches

Error in estimating the number of filled caches:



Case Studies

Three case studies:

- Stalkerware
- Contract Cheating
- Typo Squatting

Previously, all were hard to measure – little data available about prevalence.

Case Study #1: Stalkerware

Stalkerware: spyware used in IPV situations (Intimate Partner Violence)

- Monitors location
- Records all communication
- Can hide its presence

24 apps

- 6 dual use: Usually marketed for parental control or employee surveillance.
- 16 overt: "Undetectable," can be marketed explicitly for spying on intimate partner.

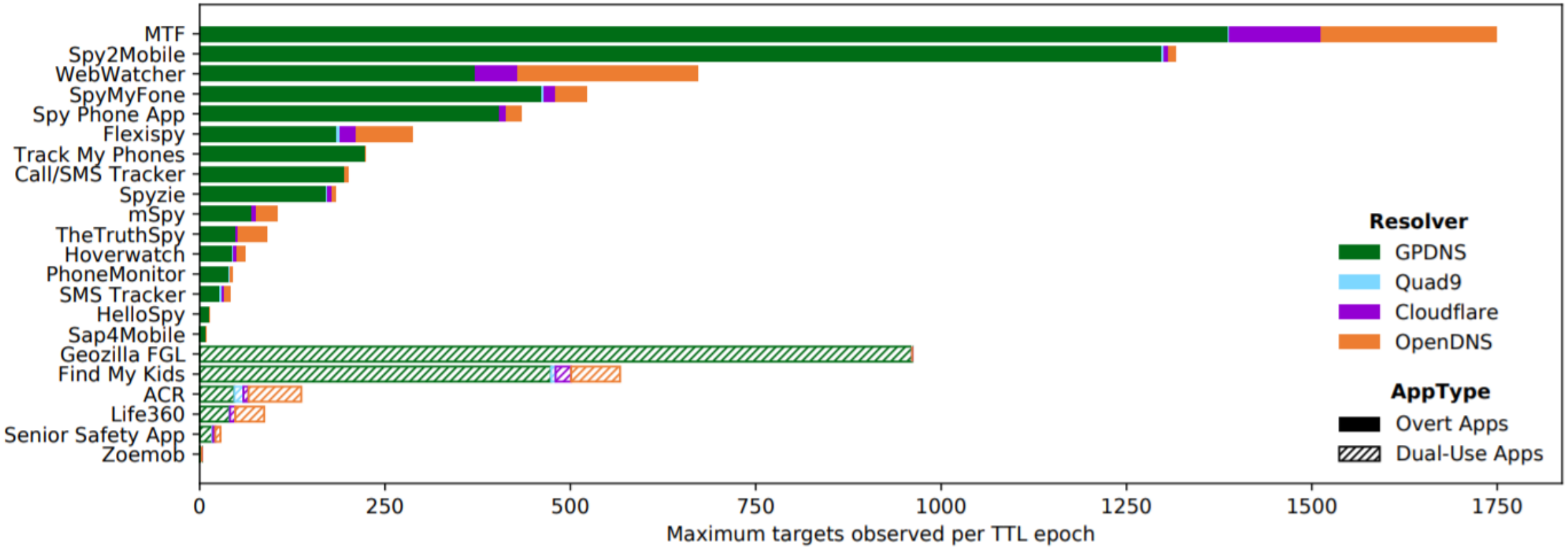
Prior work has found little to no evidence of overt apps in clinical settings. So are they being used at all?

From Counting Caches to Counting Devices

- Apps often makes DNS requests automatically, at regular intervals.
- Operating systems cache DNS responses – we assume devices make queries only once per TTL epoch.

$$\text{Devices with stalkerware installed} = \frac{\text{Filled Caches}}{\text{App Request Rate}}$$

Observed Stalkerware Targets

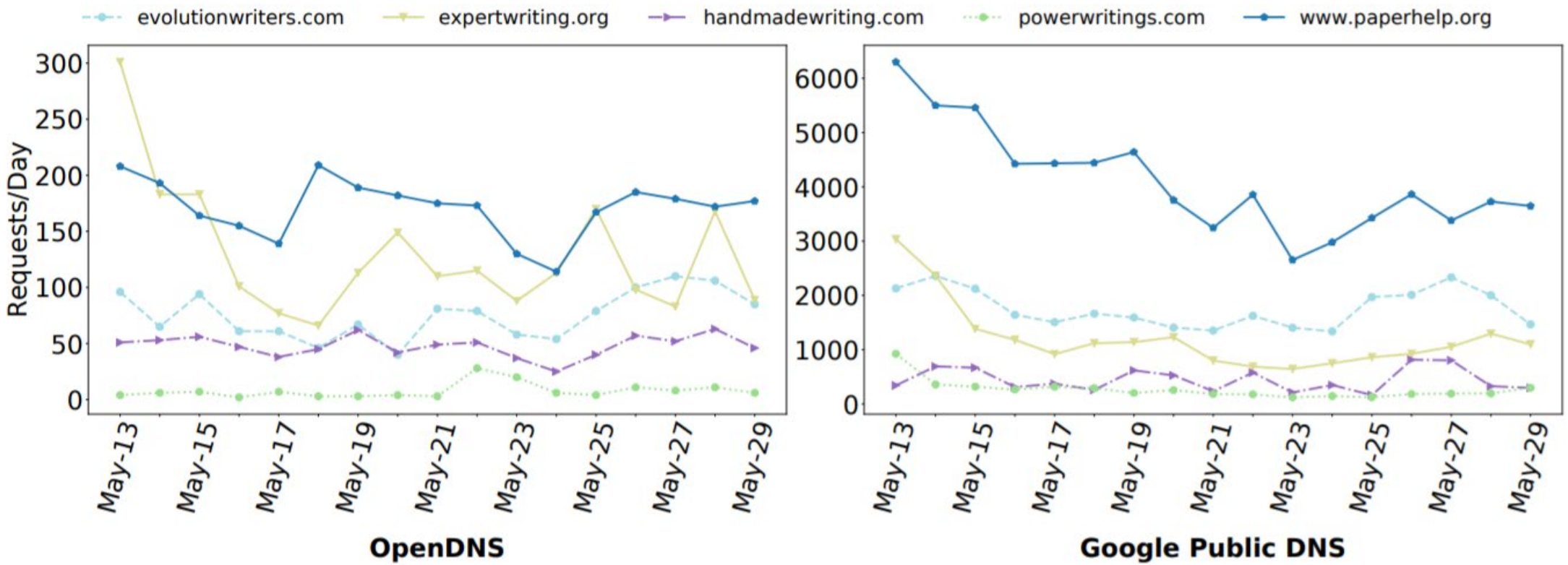


At least 5,700 people are targeted by overt stalkerware in the U.S. today.

Case Study #2: Contract Cheating

- Contract cheating is the new plagiarism!
- Students hire services to complete homework, projects, even entire classes for them
- Hard to detect – original content, can't be found with plagiarism checkers

Observed Contract Cheating



Some services decrease over time:
schools letting out for summer break?

Conclusion

- Public DNS resolvers enable privacy-preserving cache snooping
- We model the cache architecture of four public resolvers
- We present Trufflehunter, an open-source tool for measuring domain popularity
- We find non-trivial lower bounds of the popularity of previously under-studied Internet phenomena like stalkerware.

https://github.com/ucsdsysnet/trufflehunter