IETF®

*Internet Engineering Task Force 110 Meeting*
*Measurement and Analysis for Protocols Research Group*

# Assessing the Privacy Benefits of Domain Name Encryption

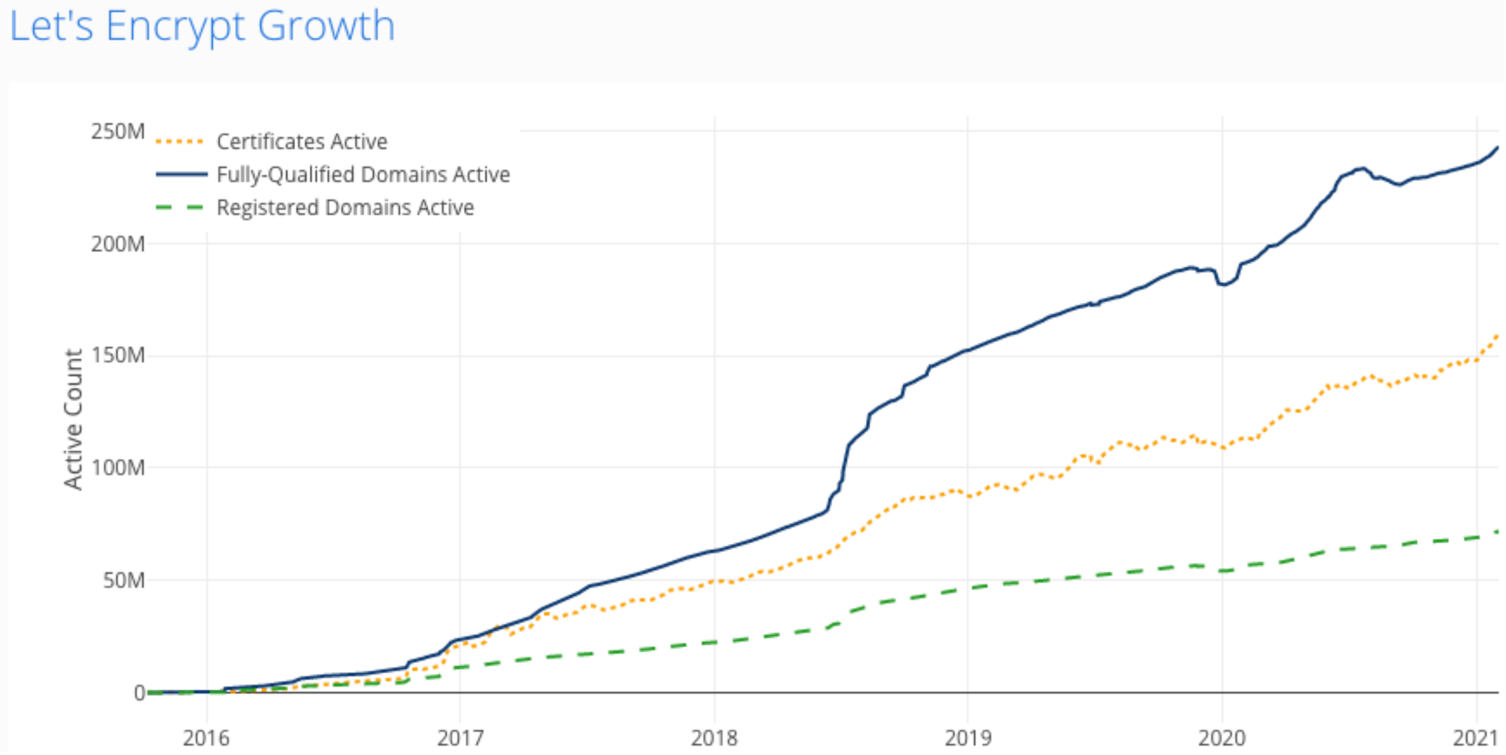
Nguyen Phong Hoang, Arian Akhavan Niaki,
Nikita Borisov, Phillipa Gill, Michalis Polychronakis

Stony Brook University
The State University of New York

UMass
Amherst

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

# Internet traffic encryption is on the rise



# Domain names still reveal semantic info

- Amazon.com, Walmart.com, Ebay.com

→ online shopping activities

- HIV.gov , Cancer.gov

→ health condition

- Islamicity.org, Quran.com

→ religion

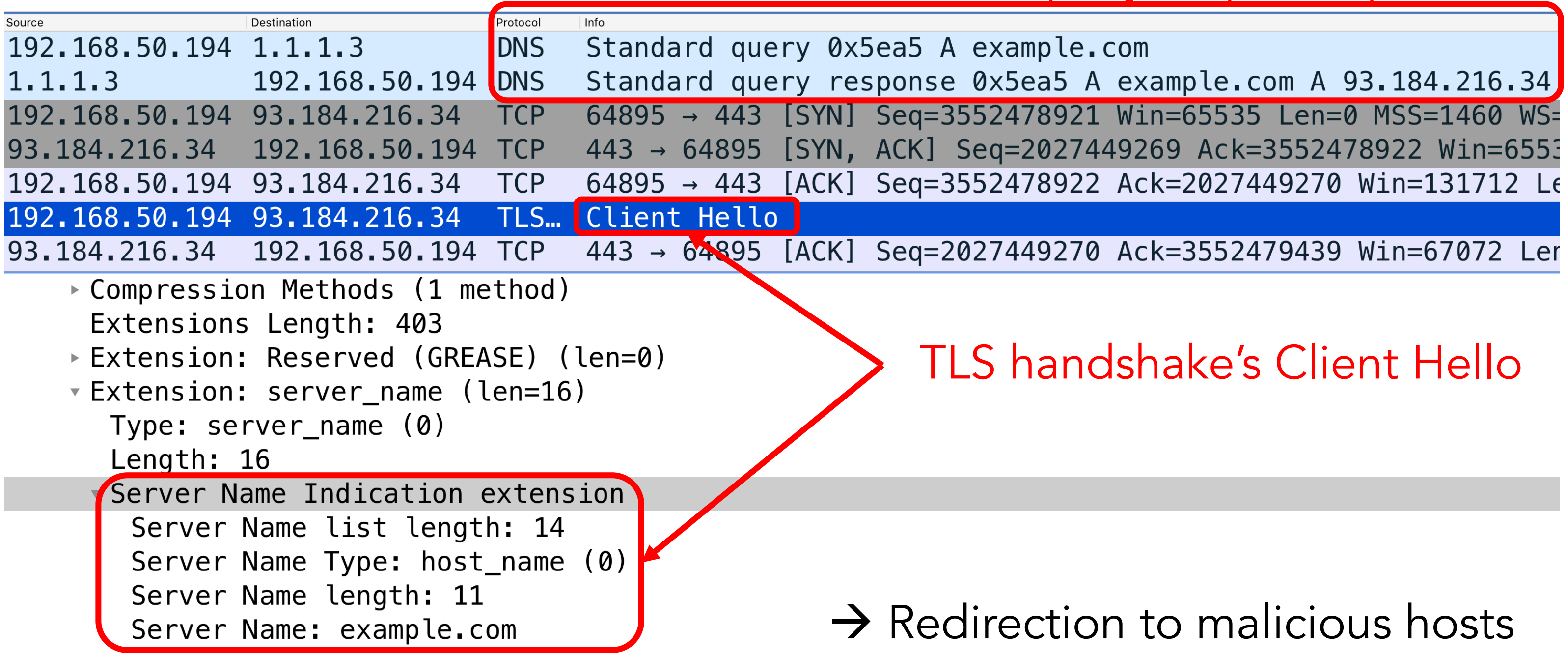
- LGBT.foundation, Gaycenter.org

→ gender identity

- Xvideos.com, Pornhub.com

→ sexual habits

# Plaintext domain name on the wire



DNS query/response packets

```
Source          Destination     Protocol  Info
192.168.50.194  1.1.1.3         DNS       Standard query 0x5ea5 A example.com
1.1.1.3         192.168.50.194  DNS       Standard query response 0x5ea5 A example.com A 93.184.216.34
192.168.50.194  93.184.216.34   TCP       64895 → 443 [SYN] Seq=3552478921 Win=65535 Len=0 MSS=1460 WS=
93.184.216.34   192.168.50.194  TCP       443 → 64895 [SYN, ACK] Seq=2027449269 Ack=3552478922 Win=6553
192.168.50.194  93.184.216.34   TCP       64895 → 443 [ACK] Seq=3552478922 Ack=2027449270 Win=131712 Le
192.168.50.194  93.184.216.34   TLS…      Client Hello
93.184.216.34   192.168.50.194  TCP       443 → 64895 [ACK] Seq=2027449270 Ack=3552479439 Win=67072 Len
```

```
    ▸ Compression Methods (1 method)
      Extensions Length: 403
    ▸ Extension: Reserved (GREASE) (len=0)
    ▾ Extension: server_name (len=16)
        Type: server_name (0)
        Length: 16
      ▾ Server Name Indication extension
          Server Name list length: 14
          Server Name Type: host_name (0)
          Server Name length: 11
          Server Name: example.com
```

TLS handshake's Client Hello

→ Redirection to malicious hosts

→ Censorship

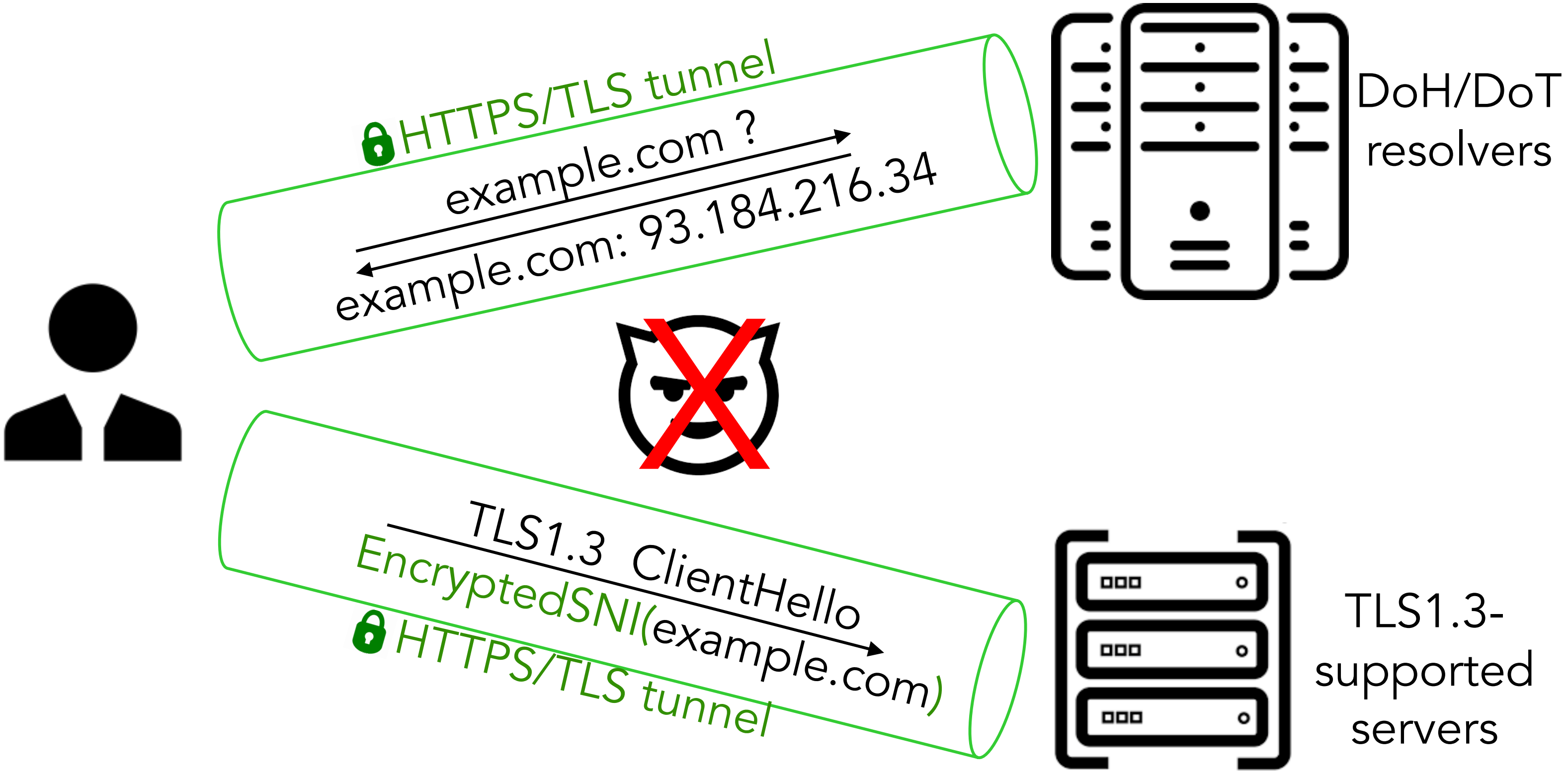
# Outline

- Introduction
  - Domain name encryption
  - Research motivation
- Measurement methodology
- Privacy benefit analysis
  - Domain co-hosting
  - Dynamics of domain-to-IP mapping
- Discussion & conclusion

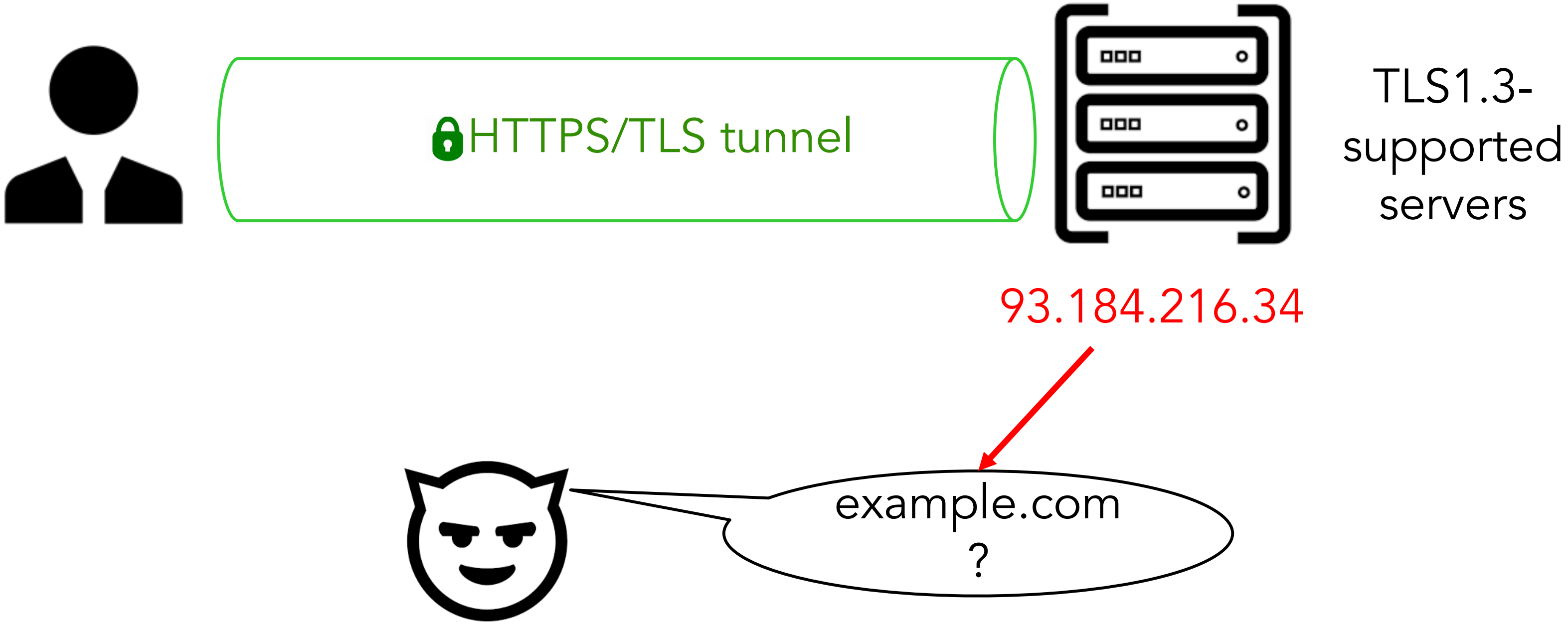
# Domain encryption: DoH/DoT and ESNI

- **DoT:** DNS queries and responses are encrypted and wrapped through the Transport Layer Security protocol (RFC7858)
- **DoH:** DNS resolution is performed over HTTPS, inheriting all security benefits of the HTTPS protocol (RFC8484)
- **ESNI**: Starting with TLS1.3, the Server Name Indication extension in the Client Hello message during the TLS handshake can be encrypted (RFC8744)

# Domain encryption: DoH/DoT and ESNI



# Domain name encryption



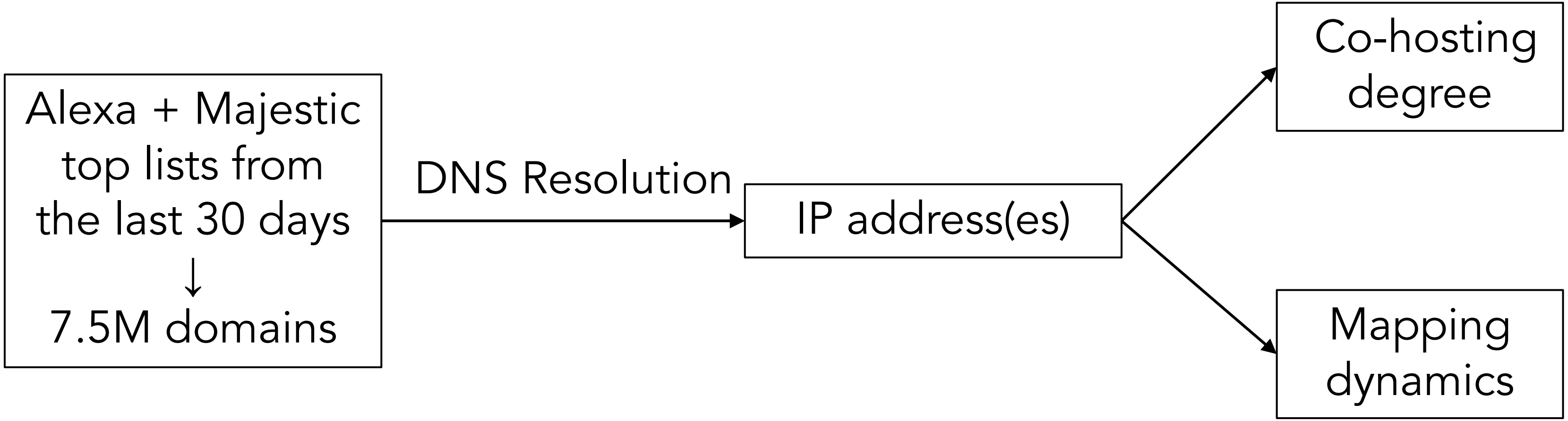
# Motivation

Given that destination IP addresses are still visible to on-path observers, we're interested in quantifying the potential improvement to user privacy that a full deployment of DoH/DoT and ESNI would achieve in practice

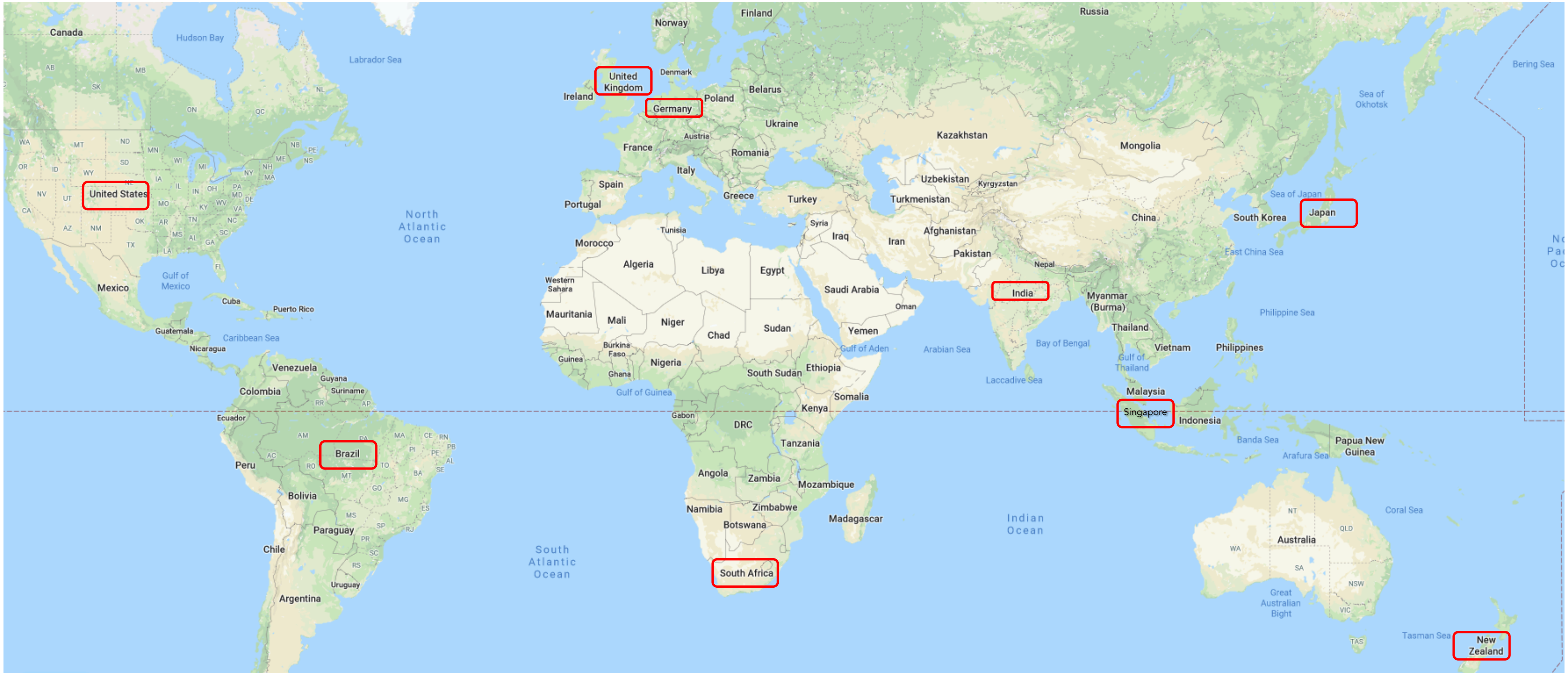
The extent to which domain inference can be made depends on:

- Whether one or many domains are hosted on a given IP address
- The stability of the mapping of a domain and its IP address(es)

# Experiment setup

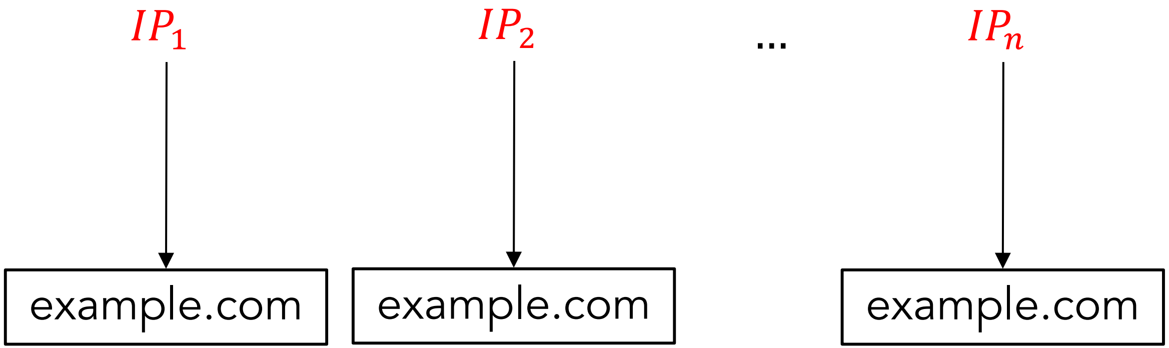


# Measurement location and duration



Measurement duration: 2 months

# Single-hosted domains

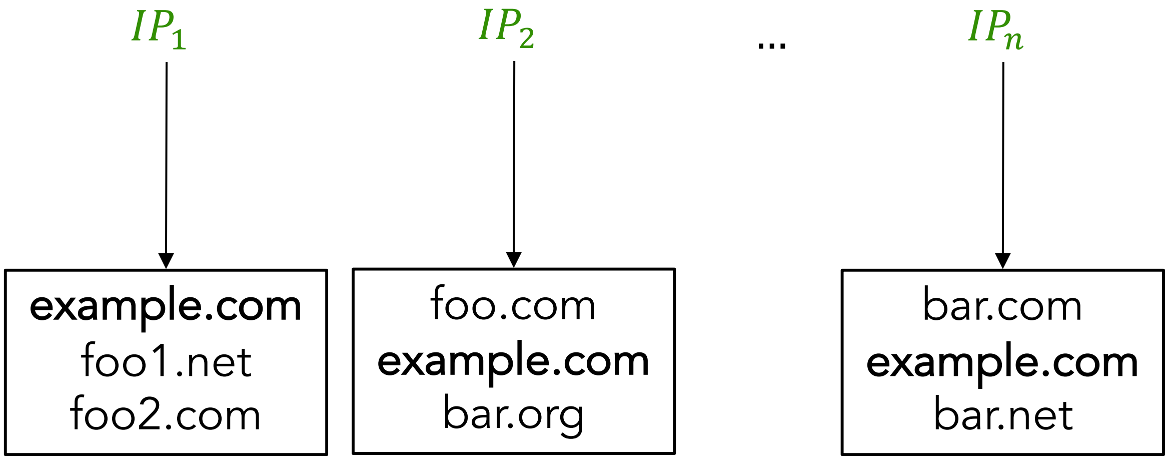


$$k_{IP_1} = k_{IP_2} = k_{IP_n} = 1$$

$$k_{example.com} = 1$$

→ Privacy-detrimental

# Multi-hosted domains



$IP_1$ $IP_2$ … $IP_n$

example.com
foo1.net
foo2.com

foo.com
example.com
bar.org

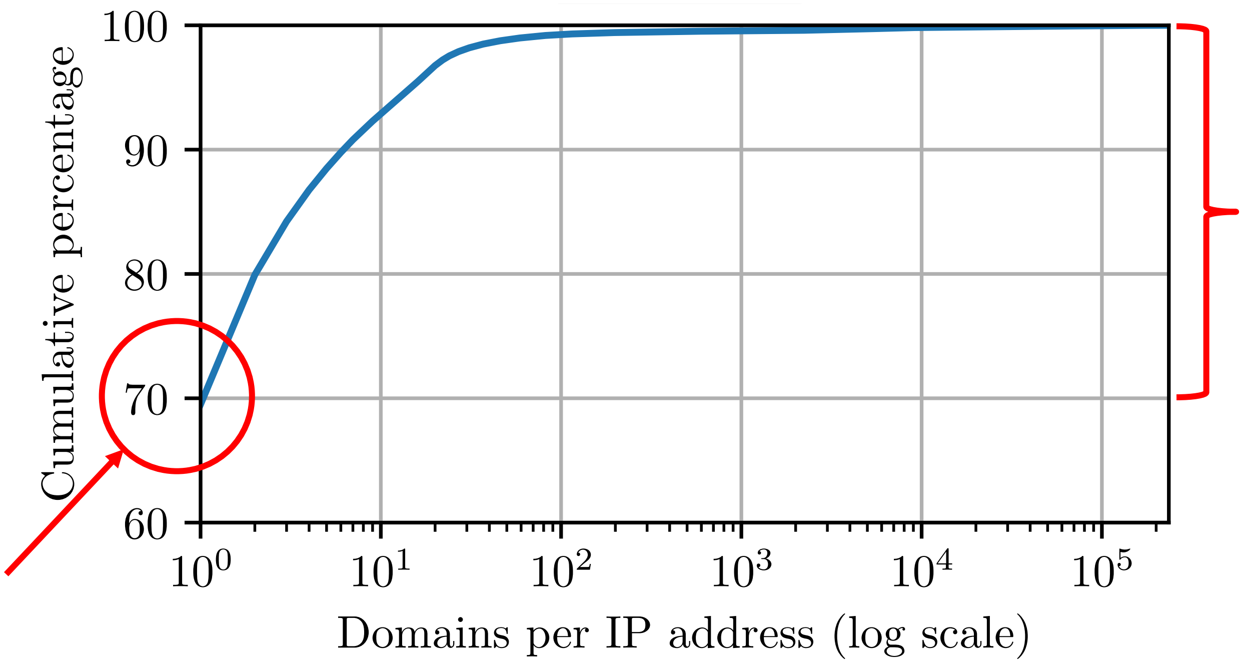
bar.com
example.com
bar.net

$$k_{IP_1} = k_{IP_2} = k_{IP_n} = 3$$

$$k_{example.com} = \text{median}(k_{IP_1}, k_{IP_2}, \ldots, k_{IP_n}) = 3$$

→ Privacy-beneficial

# Co-hosting degree as % of all IPs

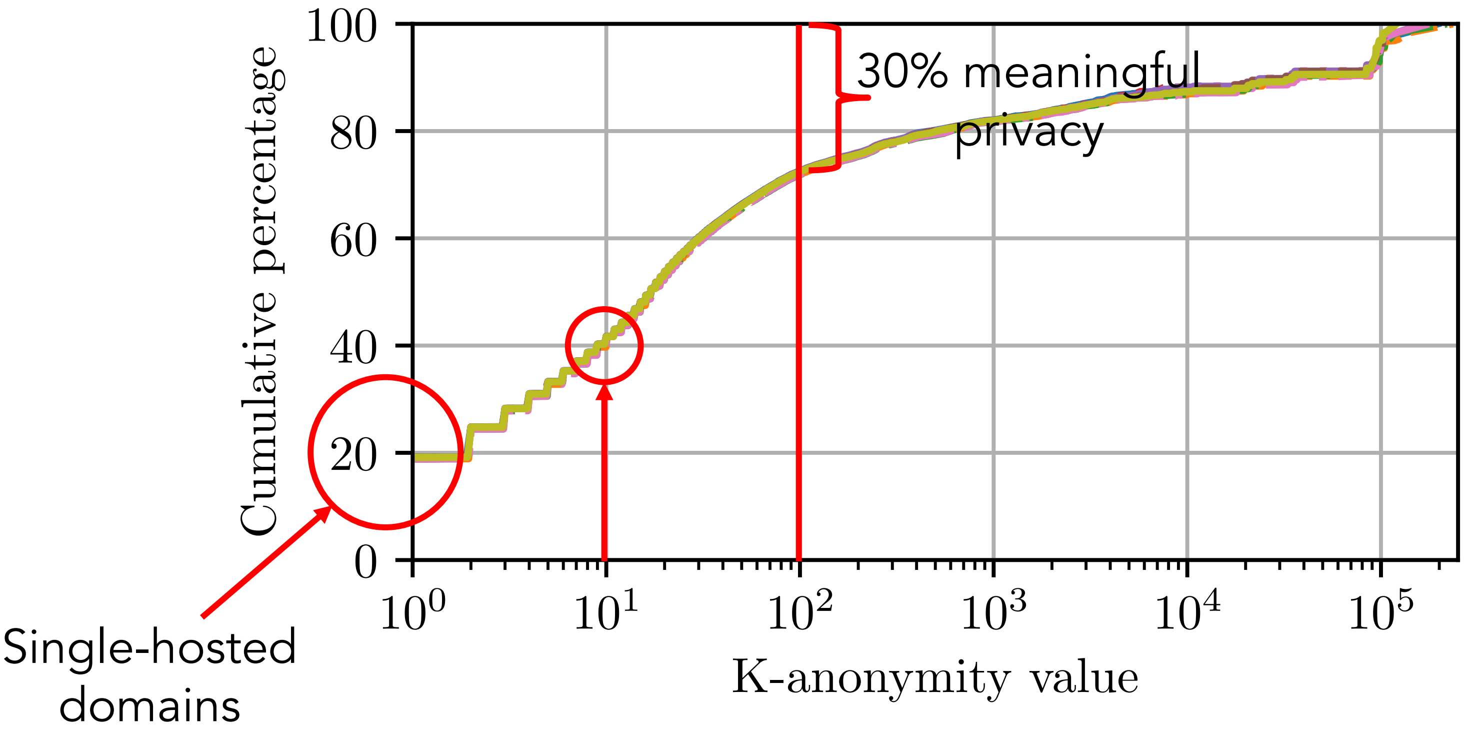


6.1M (81%) domains

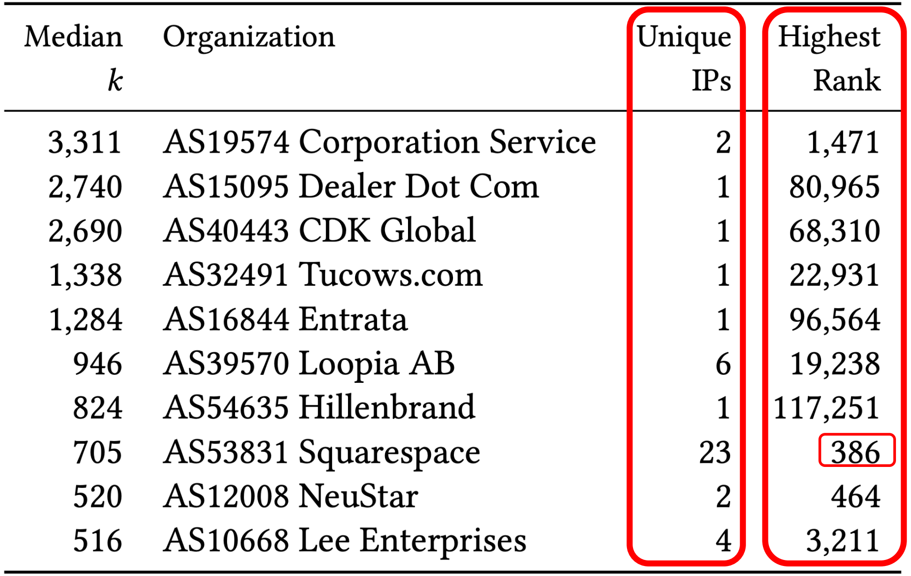
1.4M (19%) domains

Of the 2.2M IP addresses observed, 70% host only one domain

# Co-hosting degree as % of domains



# Top providers with the highest k per IP

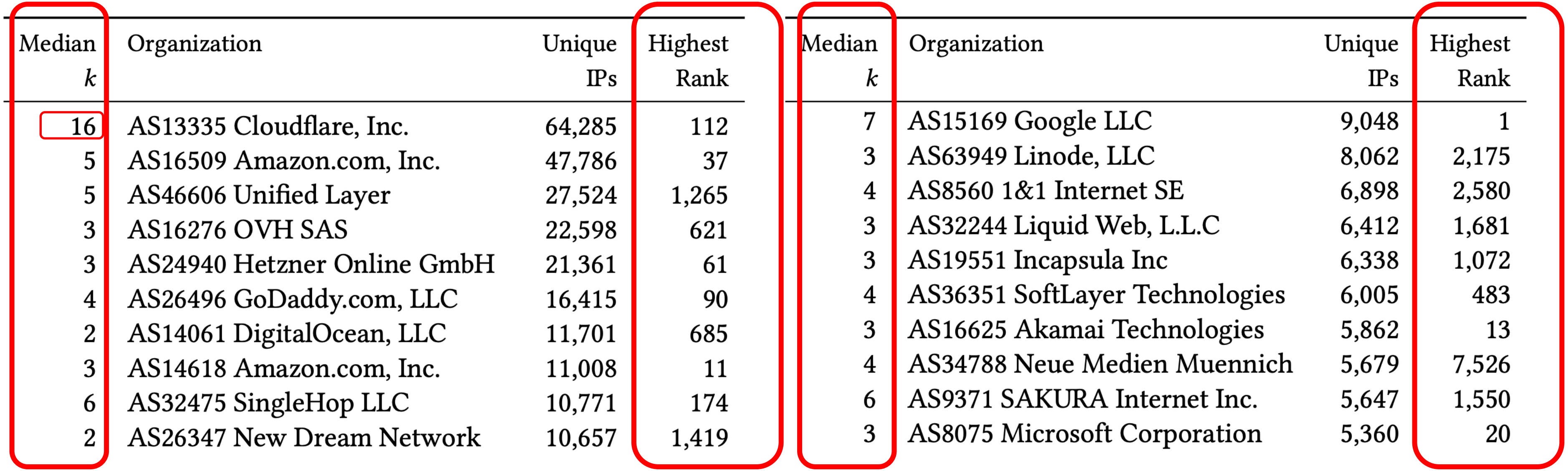


| Median<br>*k* | Organization | Unique<br>IPs | Highest<br>Rank |
|---:|---|---:|---:|
| 3,311 | AS19574 Corporation Service | 2 | 1,471 |
| 2,740 | AS15095 Dealer Dot Com | 1 | 80,965 |
| 2,690 | AS40443 CDK Global | 1 | 68,310 |
| 1,338 | AS32491 Tucows.com | 1 | 22,931 |
| 1,284 | AS16844 Entrata | 1 | 96,564 |
| 946 | AS39570 Loopia AB | 6 | 19,238 |
| 824 | AS54635 Hillenbrand | 1 | 117,251 |
| 705 | AS53831 Squarespace | 23 | 386 |
| 520 | AS12008 NeuStar | 2 | 464 |
| 516 | AS10668 Lee Enterprises | 4 | 3,211 |

Small providers tend to co-host a large number of less popular domains

*[*]* https://bgp.he.net/.

# Top providers with the most IPs

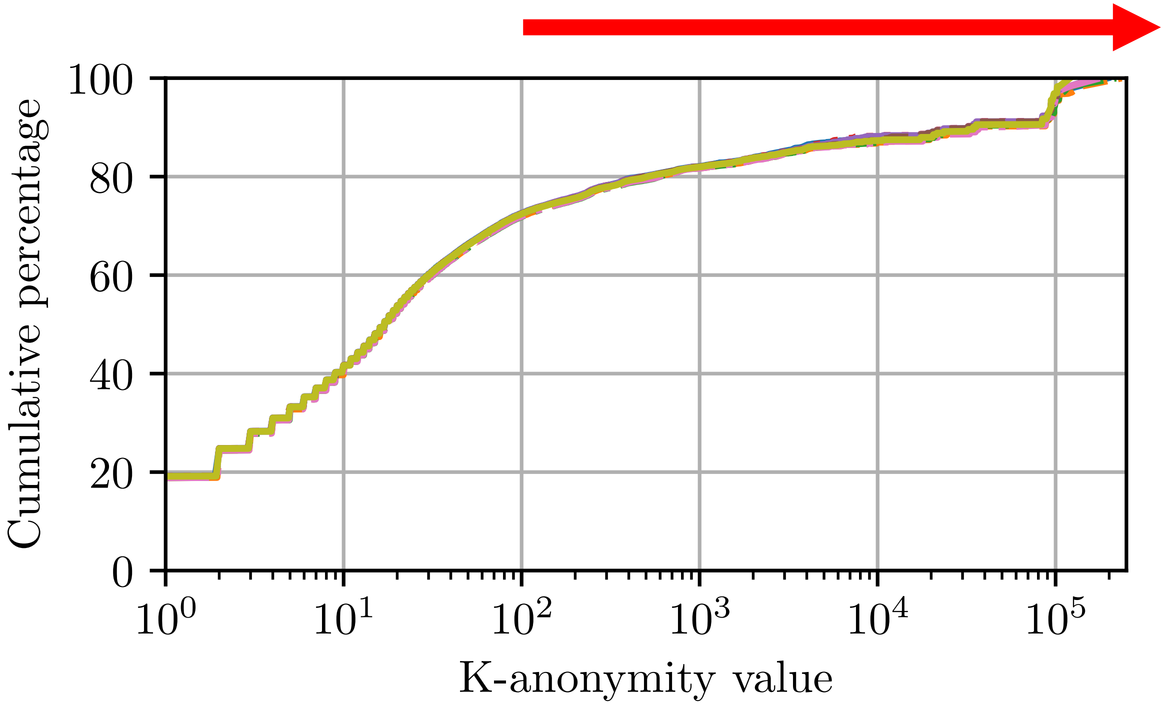


| Median<br>*k* | Organization | Unique<br>IPs | Highest<br>Rank |
|---|---|---|---|
| 16 | AS13335 Cloudflare, Inc. | 64,285 | 112 |
| 5 | AS16509 Amazon.com, Inc. | 47,786 | 37 |
| 5 | AS46606 Unified Layer | 27,524 | 1,265 |
| 3 | AS16276 OVH SAS | 22,598 | 621 |
| 3 | AS24940 Hetzner Online GmbH | 21,361 | 61 |
| 4 | AS26496 GoDaddy.com, LLC | 16,415 | 90 |
| 2 | AS14061 DigitalOcean, LLC | 11,701 | 685 |
| 3 | AS14618 Amazon.com, Inc. | 11,008 | 11 |
| 6 | AS32475 SingleHop LLC | 10,771 | 174 |
| 2 | AS26347 New Dream Network | 10,657 | 1,419 |

| Median<br>*k* | Organization | Unique<br>IPs | Highest<br>Rank |
|---|---|---|---|
| 7 | AS15169 Google LLC | 9,048 | 1 |
| 3 | AS63949 Linode, LLC | 8,062 | 2,175 |
| 4 | AS8560 1&1 Internet SE | 6,898 | 2,580 |
| 3 | AS32244 Liquid Web, L.L.C | 6,412 | 1,681 |
| 3 | AS19551 Incapsula Inc | 6,338 | 1,072 |
| 4 | AS36351 SoftLayer Technologies | 6,005 | 483 |
| 3 | AS16625 Akamai Technologies | 5,862 | 13 |
| 4 | AS34788 Neue Medien Muennich | 5,679 | 7,526 |
| 6 | AS9371 SAKURA Internet Inc. | 5,647 | 1,550 |
| 3 | AS8075 Microsoft Corporation | 5,360 | 20 |

Major providers host more popular domains,
while having a much lower co-hosting degree

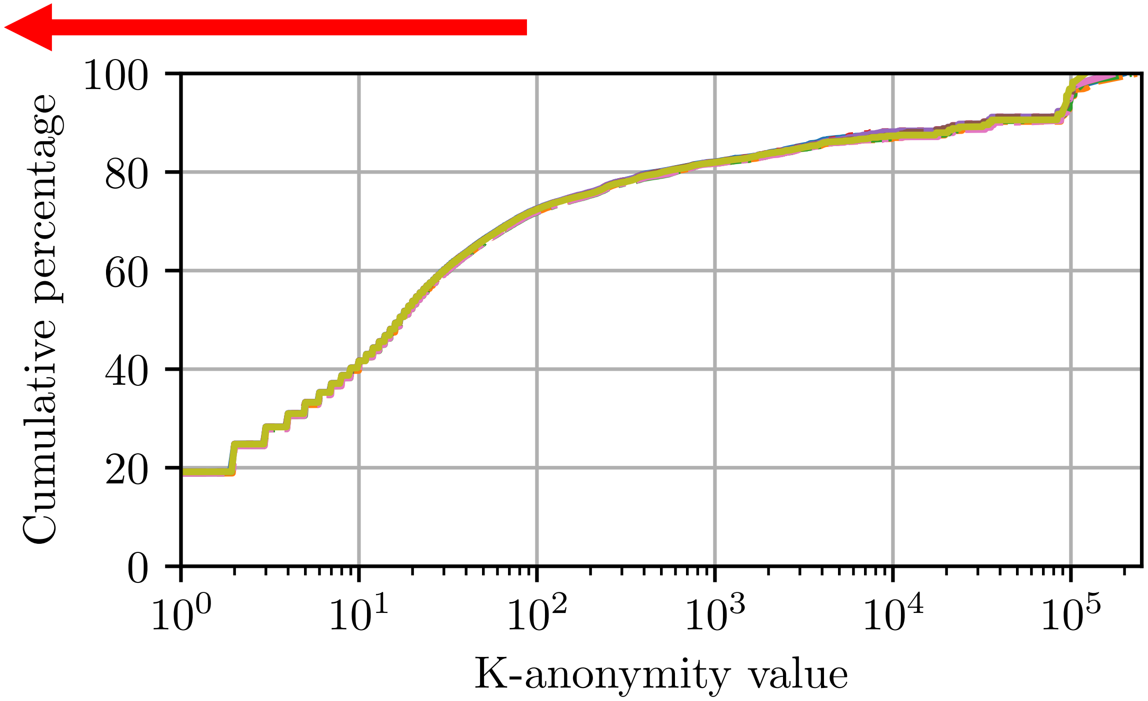
# Two ends of the privacy spectrum



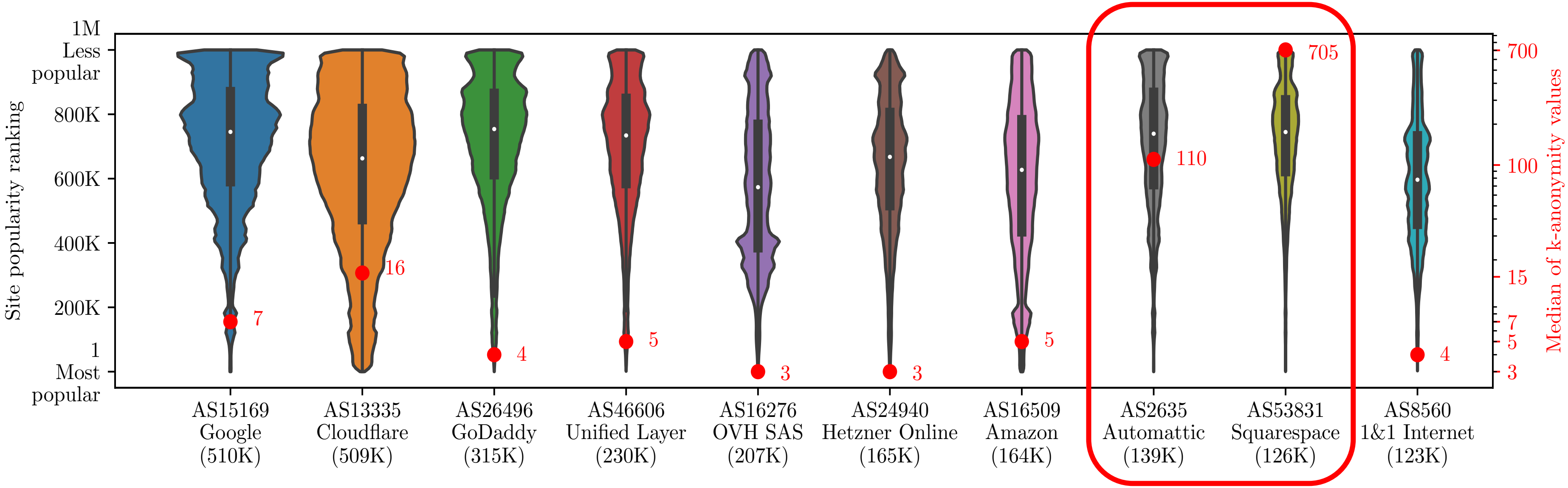
Less popular domains are hosted on smaller providers with a handful of IP addresses, benefiting from a higher k

# Two ends of the privacy spectrum

More <u>popular domains</u> are hosted on providers with a much larger pool of IP addresses, suffering from a <u>lower k</u>



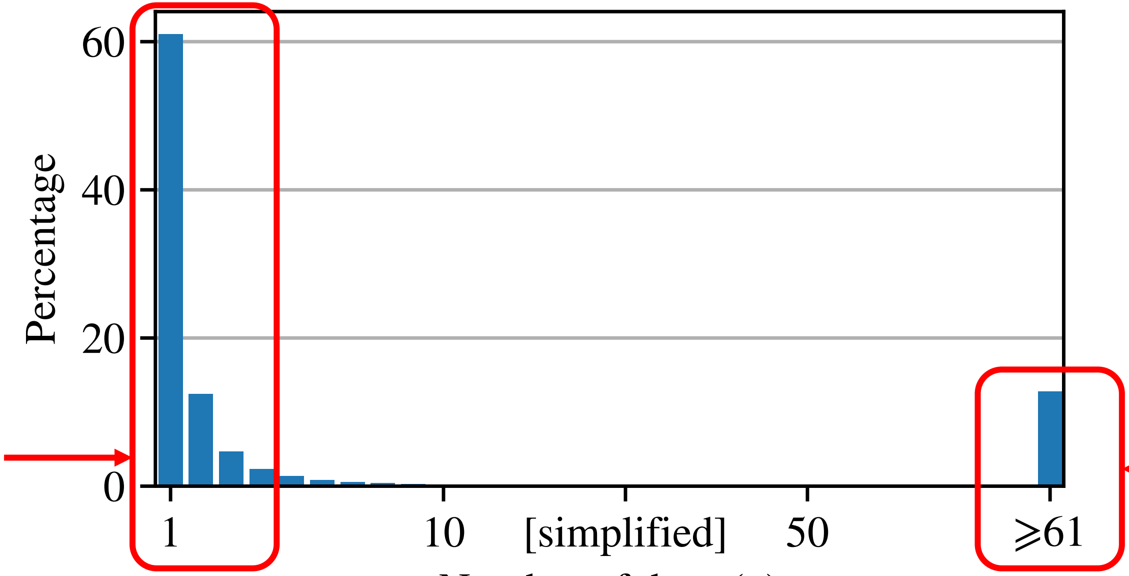
# Top providers that host most domains



- Squarespace is home to a large number of websites thanks to its pre-built template service, making it easier for anyone to build their own website
- Automattic is well-known for its WordPress service

# Dynamics of domain-to-IP mappings

2.6M domains → 22.7M domain-to-IP mappings



0.2M (7.7%) domains

2.4M (92%) domains

Dynamic 80%

Stable 13%

→ Most domains are hosted on static IP addresses

# Summary

Regardless of the increasing trend of web co-location [*], domain name encryption cannot provide meaningful privacy benefits given the current degree of domain co-hosting because the IP address information is still visible to any on-path observers and can be used to infer the domains being visited

*[*] The Web is Still Small After More Than a Decade. SIGCOMM Computer Communication Review 2020.*

# Recommendations

- **The full domain name confidentiality** must be preserved on both DNS and TLS channels; otherwise, neither technology can provide any actual privacy benefit if deployed individually
- **Domain owners** can seek providers that offer an increased co-hosting ratio per IP address and/or highly dynamic domain-IP mappings
- **Hosting providers** can help to increase the co-hosting degree by grouping more websites under the same IP and dynamically rotate domain-IP mappings to further improve privacy

# Thank you for your attention

We have made our dataset available at

https://bit.ly/DomainNameEncryptionPrivacy

*nghoang@cs.stonybrook.edu*