Assessing the Privacy Benefits of Domain Name Encryption

Nguyen Phong Hoang, Arian Akhavan Niaki, Nikita Borisov, Phillipa Gill, Michalis Polychronakis
Internet traffic encryption is on the rise

Let's Encrypt Growth

![Graph showing the growth of internet traffic encryption from 2016 to 2021.](image)
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

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](https://tools.ietf.org/html/rfc7858))

- **DoH**: DNS resolution is performed over HTTPS, inheriting all security benefits of the HTTPS protocol ([RFC8484](https://tools.ietf.org/html/rfc8484))

- **ESNI**: Starting with TLS1.3, the Server Name Indication extension in the Client Hello message during the TLS handshake can be encrypted ([RFC8744](https://tools.ietf.org/html/rfc8744))
Domain encryption: DoH/DoT and ESNI

- **HTTPS/TLS tunnel**
  - `example.com`?
  - `example.com: 93.184.216.34`

- **TLS1.3 ClientHello**
  - `EncryptedSNI(example.com)`

- **TLS1.3-supported servers**

---

**DoH/DoT resolvers**
Domain name encryption

TLS1.3-supported servers

93.184.216.34

HTTPS/TLS tunnel

example.com?
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

Alexa + Majestic top lists from the last 30 days \(\downarrow\) 7.5M domains

DNS Resolution

IP address(es)

Co-hosting degree

Mapping dynamics
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

\[ k_{IP_1} = k_{IP_2} = k_{IP_n} = 3 \]

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

→ Privacy-beneficial
Of the 2.2M IP addresses observed, 70% host only one domain.
Co-hosting degree as % of domains

Cumulative percentage vs. K-anonymity value graph showing:

- 30% meaningful privacy
- Single-hosted domains

Introduction
Methodology
Data Analysis
Discussion & Conclusion
Top providers with the highest k per IP

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

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

[1] https://bgp.he.net/
Top providers with the most IPs

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

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

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 popular domains are hosted on providers with a much larger pool of IP addresses, suffering from a lower $k$. 
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 $\rightarrow$ 22.7M domain-to-IP mappings

- 2.4M (92%) domains
- 0.2M (7.7%) domains

$\rightarrow$ Most domains are hosted on static IP addresses

Number of days (n)

Percentage

0 10 20 30 40 50 60

Dynamic 80%

Stable 13%

$\geq 61$
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.

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


ngohoang@cs.stonybrook.edu