What Can You Learn from an IP?

Simran Patil and Nikita Borisov

University of Illinois at Urbana-Champaign

@SimranPatil25 @nikitab
In the beginning...

GET /~nikitab/ HTTP/1.1
Host: geocities.com
...

HTTP/1.1 200 OK
...
<br> <blink>this page is under construction</blink>

http://geocities.com /~nikitab/
under construction
Today

DNS query

A? irtf.org

irtf.org A 4.31.198.44

ClientHello
... SNI irtf.org

Server Certificate
... CN=irtf.org

TLS handshake

GET /anrw/2019/ HTTP/1.1
Host: irtf.org

HTTP/1.1 200 OK
...
<title>ANRW’19</title>

TLS encrypted

https://irtf.org/???
???
Soon?

DNS query

A? irtf.org

irtf.org A 4.31.198.44

TLS handshake

ClientHello
... SNI irtf.org

Server Certificate
... CN=irtf.org

GET /anrw/2019/ HTTP/1.1
Host: irtf.org
...

HTTP/1.1 200 OK
...
<title>ANRW’19</title>

DNS-over-HTTPS/TLS

ESNI

TLS1.3

4.31.198.44

ANRW’19

S. Patil & N. Borisov, "What Can You Learn from an IP?"
What can you learn from a domain name?

- drugrehab.ca
- vim.org
- whatisabrony.com
- dailystormer.name
- foxnews.com
- lymphoma.ca
- anime-expo.org
- aljazeera.com
- furrycons.com
- www.lgbtcenters.org
- nickleback.com
- www.oshawamosque.com
- montrealcathedral.ca
Methodology

Alexa global top 1000000 → MIDA → Page resources: URLs, domains, types → zdns → domains => IP address => rDNS

- 944,094 sites
- 90,514,000 objects
- 1,819,087 domains
- 1,795,506 resolved
- 741,049 IPs
rDNS

Public Suffix List (PSL) match:
server1.facebook.com =~ facebook.com
Domains and IPs

Average degree: 1.46

Average in-degree: 3.14
IP Anonymity Set

Average degree: 1.46
Average in-degree: 3.14
IP Anonymity Sets

- 47.6% IPs have an anonymity set of 1
- Largest anonymity set has 16,050 domains
Site-unique IPs

E.g., 74.125.132.154 has an anonymity set of 1—stats.g.doubleclick.net—but is seen on over 10% of all the sites in our data set!
Site-unique IPs

- 68% of IPs in our set are site-unique
- 43% of sites use at least 1 resource that maps to a site-unique IP
- For 39.5% of sites, the front page maps to a site-unique IP
## Page Load Fingerprints

<table>
<thead>
<tr>
<th>Resource Name</th>
<th>URL</th>
<th>Type</th>
<th>IP Address</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>queen-elizabeth-montreal</td>
<td><a href="http://www.fairmont.com">www.fairmont.com</a></td>
<td>document</td>
<td>23.64.109.196:443</td>
<td></td>
</tr>
<tr>
<td>font-prod.js</td>
<td>cloud.moovweb.net</td>
<td>js</td>
<td>192.33.31.70:443</td>
<td></td>
</tr>
<tr>
<td>font.css</td>
<td><a href="http://www.fairmont.com">www.fairmont.com</a></td>
<td>css</td>
<td>23.64.109.196:443</td>
<td></td>
</tr>
<tr>
<td>fairmont.min.css</td>
<td>d1r80pgesju2u.cloud...</td>
<td>css</td>
<td>99.84.112.4:443</td>
<td></td>
</tr>
<tr>
<td>tripadvisor.min.css</td>
<td>d1r80pgesju2u.cloud...</td>
<td>css</td>
<td>99.84.112.4:443</td>
<td></td>
</tr>
<tr>
<td>3colbigbanner.min.css</td>
<td><a href="http://www.fairmont.com">www.fairmont.com</a></td>
<td>css</td>
<td>23.64.109.196:443</td>
<td></td>
</tr>
<tr>
<td>s_code.js</td>
<td>d1r80pgesju2u.cloud...</td>
<td>js</td>
<td>99.84.112.4:443</td>
<td></td>
</tr>
<tr>
<td>accorhotelsconnect.js</td>
<td>secure.accorhotels.ws</td>
<td>js</td>
<td>193.200.231.133:443</td>
<td></td>
</tr>
<tr>
<td>WebResource.axd</td>
<td><a href="http://www.fairmont.com">www.fairmont.com</a></td>
<td>js</td>
<td>23.64.109.196:443</td>
<td></td>
</tr>
<tr>
<td>593e534a-7cf8-4956-b5a9-4c8da07c...</td>
<td><a href="http://www.fairmont.com">www.fairmont.com</a></td>
<td>jpg</td>
<td>23.64.109.196:443</td>
<td></td>
</tr>
</tbody>
</table>
Site IP sets

95.7% sites have a unique IP set

Cluster of 903 sites has same IP set
What about CDNs?

• Many CDNs **could** use same IP address for all sites but **don’t**
  • Ported IP space
  • Connections w/o SNI

• In our data set 200K domains are hosted by CloudFlare, using 91K IPs
  • Including 3% of the sites with a site-unique front page IP

• Randomizing or normalizing IP addresses could help
Conclusions

• DNS privacy offers limited protection
  • For web browsing
  • Against an adversary with a good prior list of sites

• In our Alexa 1M crawl dataset
  • 48% of all IPs map to a single domain
  • 68% of all IPs map to a single site
  • 43% of all sites contain a site-unique IP
  • 95% of sites have a unique IP set

• Changes to web hosting infrastructure could help
  • Normalize or randomize CDN IP addresses