Encrypted DNS → Privacy?
A Traffic Analysis Perspective

Sandra Siby, Marc Juarez, Claudia Diaz, Narseo Vallina-Rodriguez, Carmela Troncoso

IETF 105, 24 July 2019
What did we do?

Conducted a number of experiments that showed that:

- Monitoring and censorship are feasible even when DNS is encrypted.
- Current proposed EDNS0-based countermeasures are not sufficient to prevent traffic analysis attacks.
The Past

Query: google.com?

Response: 172.217.168.4

Encrypted

HTTP requests and responses

Destination Host 172.217.168.4
The Future?

**DNS-over-TLS (DoT)**
**DNS-over-HTTPS (DoH)**

**Client**

Query: google.com?

**Recursive Resolver**

Response: 172.217.168.4

**Name Servers**

Encrypted

HTTP requests and responses

**Destination Host**

172.217.168.4
Scenario

What information is available to the observer?

Size, timing, directionality, headers
Scenario

What information is available to the observer?

Size, timing, directionality, headers

Key Idea: A webpage visit can have multiple DNS queries/responses associated with it, which could be a fingerprint for identification of that webpage.
Scenario

- **Client**
- **Recursive Resolver**
- **Observer**

**DNS-over-HTTPS traffic**

**Adversary goals**
- Monitoring
- Censorship
Adversary Goal 1: Monitoring

Train a classifier on size and directionality features.

Experiment 1

- Adversary knows the complete set of webpages visited by a user.
- Which particular webpage did the user visit?
- 1,500 webpages

Experiment 2

- User can visit webpages outside of the adversary’s monitored set.
- Did the user visit a page in the monitored set?
- 5,000 webpages

~90% Precision and Recall  ~70% Precision and Recall
Adversary Goal 2: Censorship

*Censoring adversary: Identify webpages as fast as possible*

Study the uniqueness of DoH traffic when only the first \( L \) TLS records have been observed (set of 1,500 pages).
Adversary Goal 2: Censorship

Censoring adversary: Identify webpages as fast as possible

Study the uniqueness of DoH traffic when only the first \( L \) TLS records have been observed (set of 1,500 pages).

Adversary strategy: **Block on first query?**

- 4th record usually corresponds to first DoH query.

Adversary strategy: **High confidence guessing?**

- By 15th record (15% of trace), most traces are distinguishable.
Robustness of attack

Key takeaway:

Changes in the setup scenario affect, but do not stop, the attack.
Monitoring and Censorship are feasible even when DNS traffic is encrypted

Countermeasures?
EDNS0 Based Countermeasures

EDNS0: Extension mechanisms for DNS, specifies a padding option¹

Padding of DNS queries: We implemented the recommended padding strategy² on Cloudflare’s DoH client. Pad query to multiples of 128 bytes.

¹RFC7830
²RFC8467
Padding of DNS responses: Cloudflare’s resolver pads responses to multiples of 128 bytes. Recommended strategy: Pad to multiples of 468 bytes
## Our experiments

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDNS0-128</td>
<td>Cloudflare’s response padding strategy</td>
</tr>
<tr>
<td>EDNS0-468</td>
<td>Recommended response padding strategy</td>
</tr>
<tr>
<td>Constant Padding</td>
<td>Keep all TLS record sizes constant</td>
</tr>
<tr>
<td>DNS over Tor</td>
<td>Cloudflare’s DNS over Tor service</td>
</tr>
</tbody>
</table>
Results: Classifier performance

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDNS0-128</td>
<td>0.710 ± 0.005</td>
<td>0.700 ± 0.004</td>
<td>0.691 ± 0.004</td>
</tr>
<tr>
<td>EDNS0-468</td>
<td>0.452 ± 0.007</td>
<td>0.448 ± 0.006</td>
<td>0.430 ± 0.007</td>
</tr>
<tr>
<td>Constant Padding</td>
<td>0.070 ± 0.003</td>
<td>0.080 ± 0.002</td>
<td>0.066 ± 0.002</td>
</tr>
<tr>
<td>DNS over Tor</td>
<td>0.035 ± 0.004</td>
<td>0.037 ± 0.003</td>
<td>0.033 ± 0.003</td>
</tr>
</tbody>
</table>

*EDNS0 based measures do not eliminate traffic analysis attacks*
Results: Overhead

Sent + received bytes (from TLS records)
Anonymous communication as a defense

Fixed cell sizes
  • Affect size features

Repacketization
  • Affect directionality features

Clusters in confusion graph?

Pages in a cluster are misclassified as each other

Confusion graph of misclassified labels
Ongoing/Next Steps

Realistic scenarios

• Multi-tab browsing
  ▶ ~40% Precision/Recall for 0.5s interval between tabs

• Caching

Comparison with DNS over TLS

• Preliminary results with padding: ~28% Precision/Recall

Countermeasures

• Padding + repacketization measures — Can we do repacketization without using Tor?
Summary

• Surveillance and DNS-based censorship can occur even in the presence of encrypted DNS.
• Current proposed EDNS0 based countermeasures are not sufficient.
• Recommendation: Repacketization and padding


Get in touch:  sandra.siby@epfl.ch  @sansib
BACKUP
Do we even need DNS traffic analysis?

Use IP address of destination host?

Virtual hosts
CDNs

Destination hostname revealed during TLS setup

TLS 1.3
Encrypted SNI
Feature extraction

pcap file ➔ 24 -58 63 110 -92 -86 -55

TLS record sizes

Burst sizes

Single record sizes

Uni-grams: (24), (-58)…

Bi-grams: (24, -58), (-58, 63)…

Uni-grams: (24), (-58)…

Bi-grams: (24, -58), (-58, 173)…

Counts