Encrypted DNS ⟶ Privacy?

A Traffic Analysis Perspective

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

IETF 105, 24 July 2019

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

EPFL

KU LEUVEN

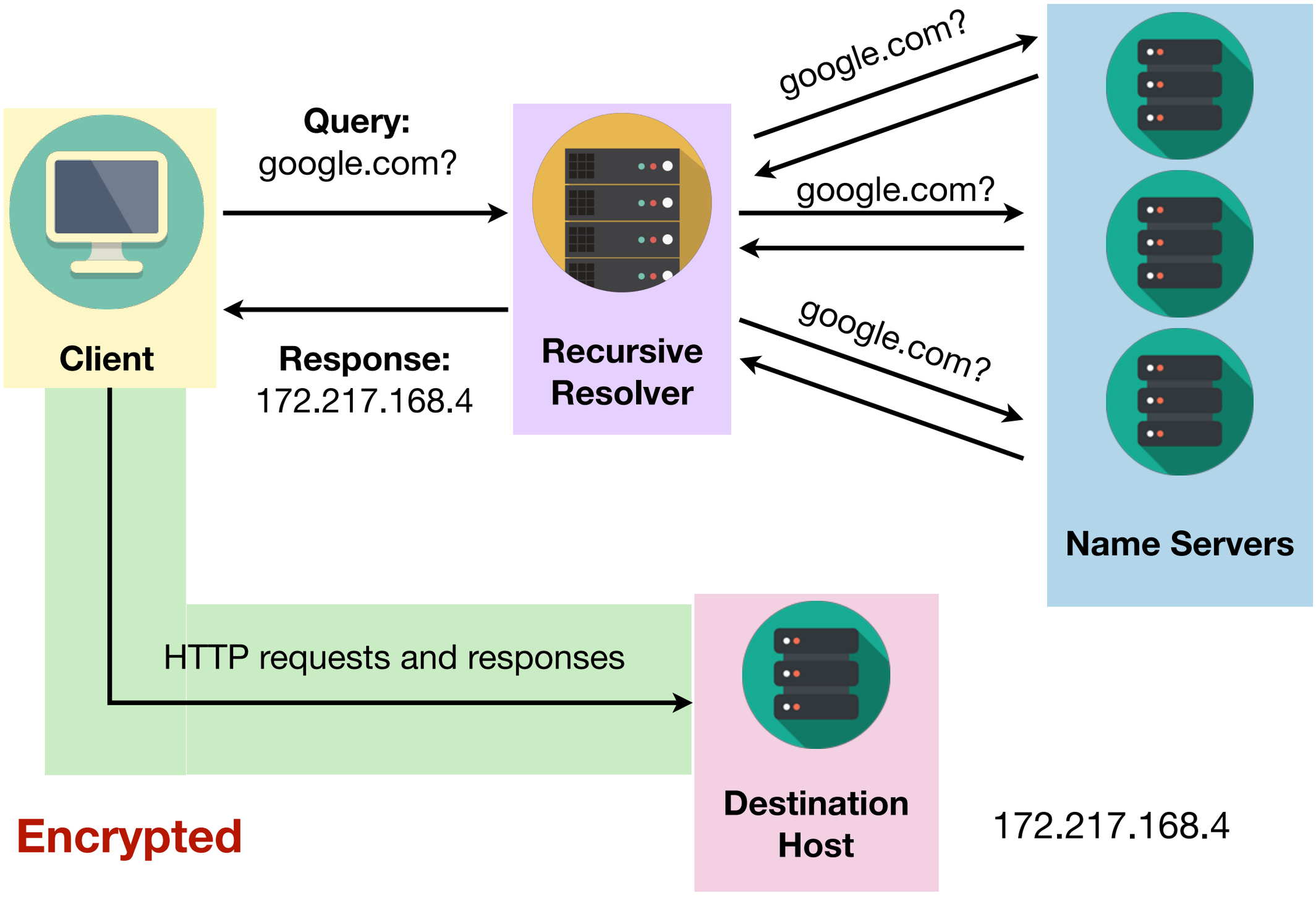
institute imdea networks

What did we do?

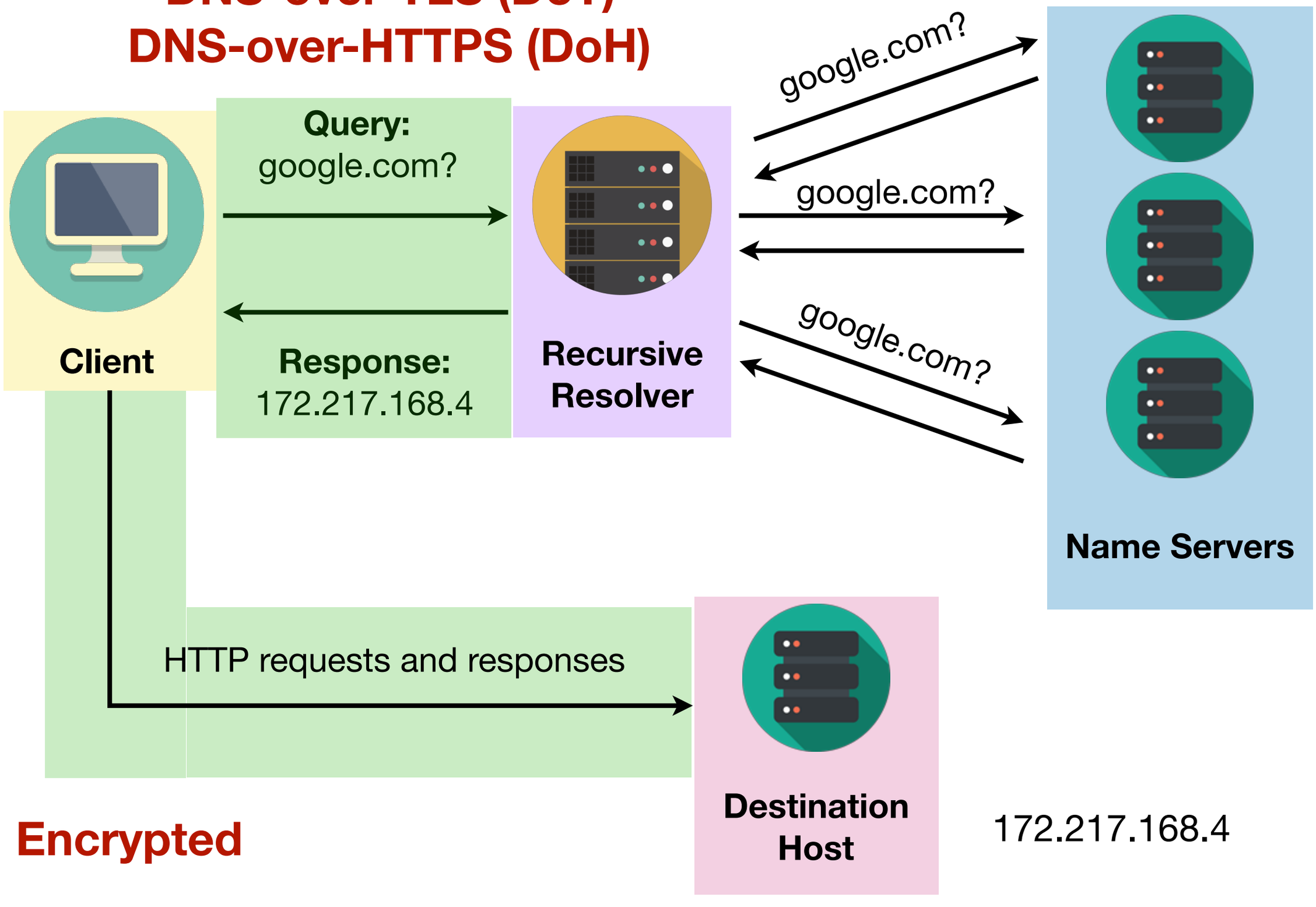
Conducted a number of experiments that showed that:

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

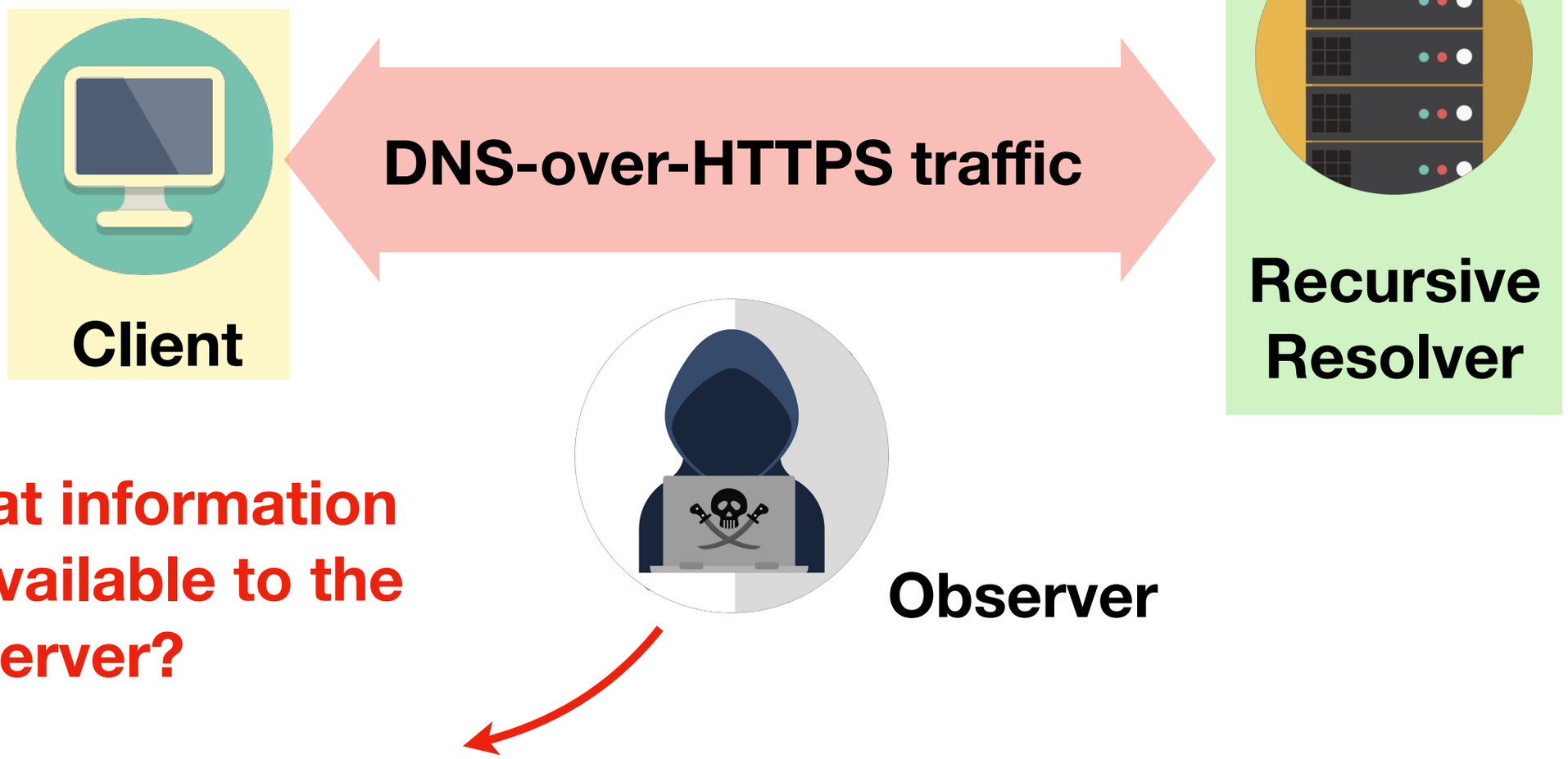
The Past


Query:
google.com?
Response:
172.217.168.4
Client
Recursive Resolver
google.com?
google.com?
google.com?
Name Servers
HTTP requests and responses
Destination Host
Encrypted
172.217.168.4

The Future?


DNS-over-TLS (DoT)
DNS-over-HTTPS (DoH)
Query:
google.com?
Response:
172.217.168.4
Client
Recursive
Resolver
google.com?
google.com?
google.com?
Name Servers
HTTP requests and responses
Destination
Host
Encrypted
172.217.168.4

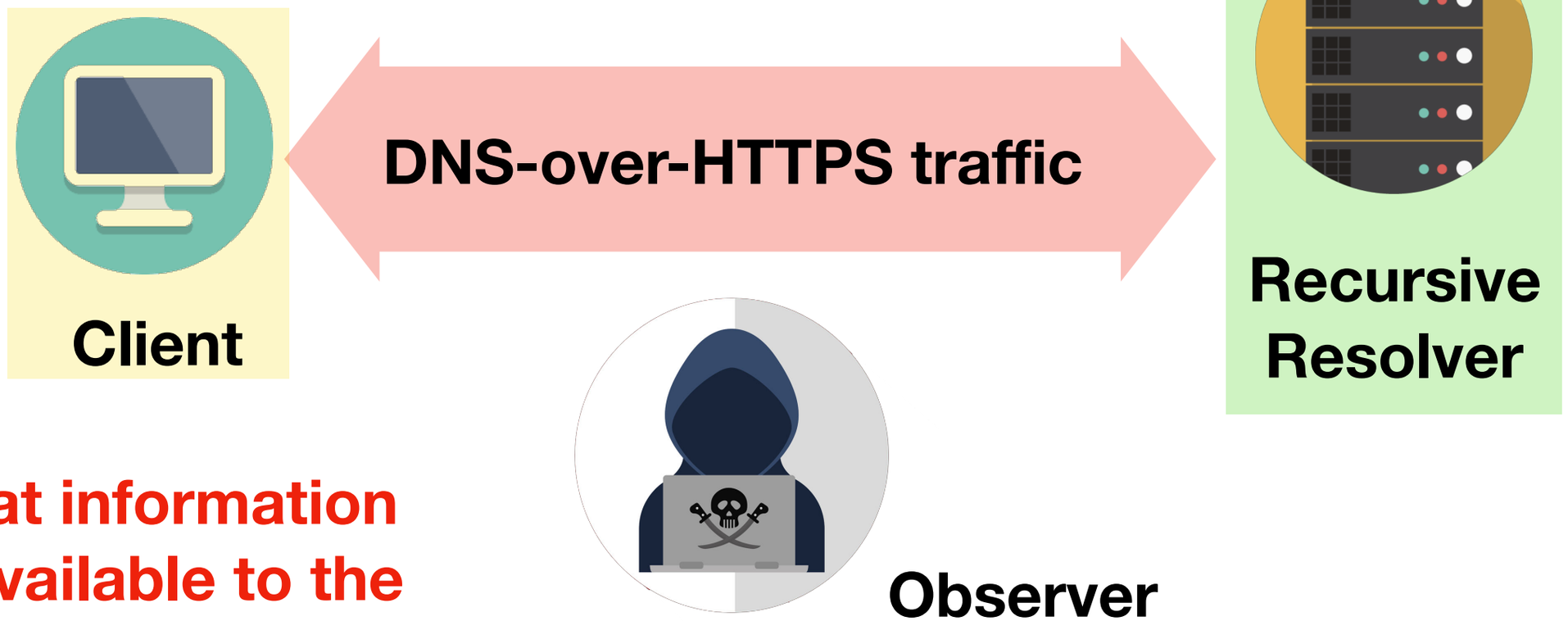
Scenario



What information is available to the observer?

Size, timing, directionality, headers

Scenario

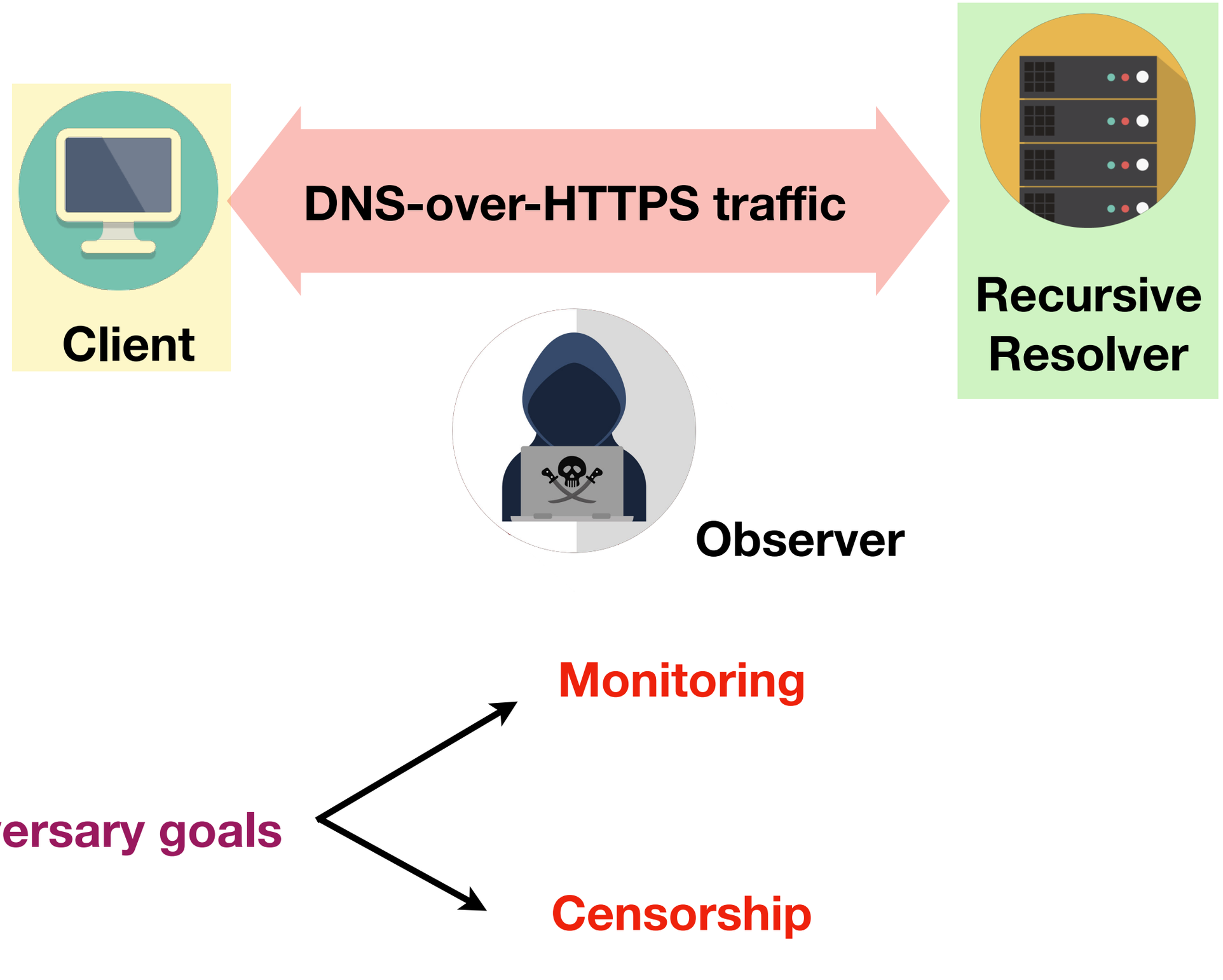


What information is available to the observer?

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


DNS-over-HTTPS traffic
Client
Recursive Resolver
Observer
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

~90% Precision and Recall

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

~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

Time

Location

Infrastructure

- Resolver
- Client
- Platform

Key takeaway:

Changes in the setup scenario affect, but do not stop, the attack.

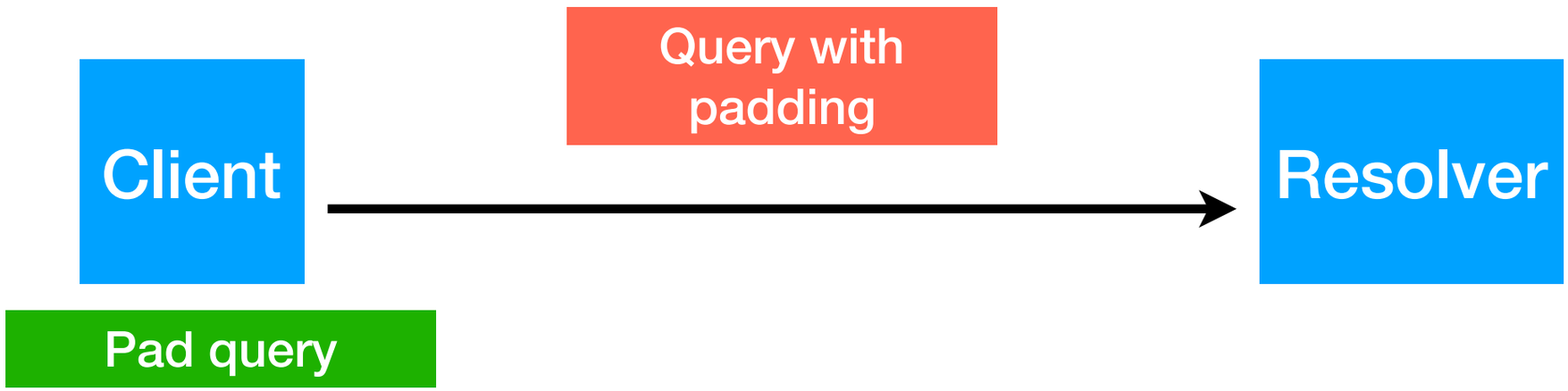
Monitoring and Censorship are feasible even when DNS traffic is encrypted

Countermeasures?

EDNS0 Based Countermeasures

EDNS0: Extension mechanisms for DNS, specifies a padding option[1]

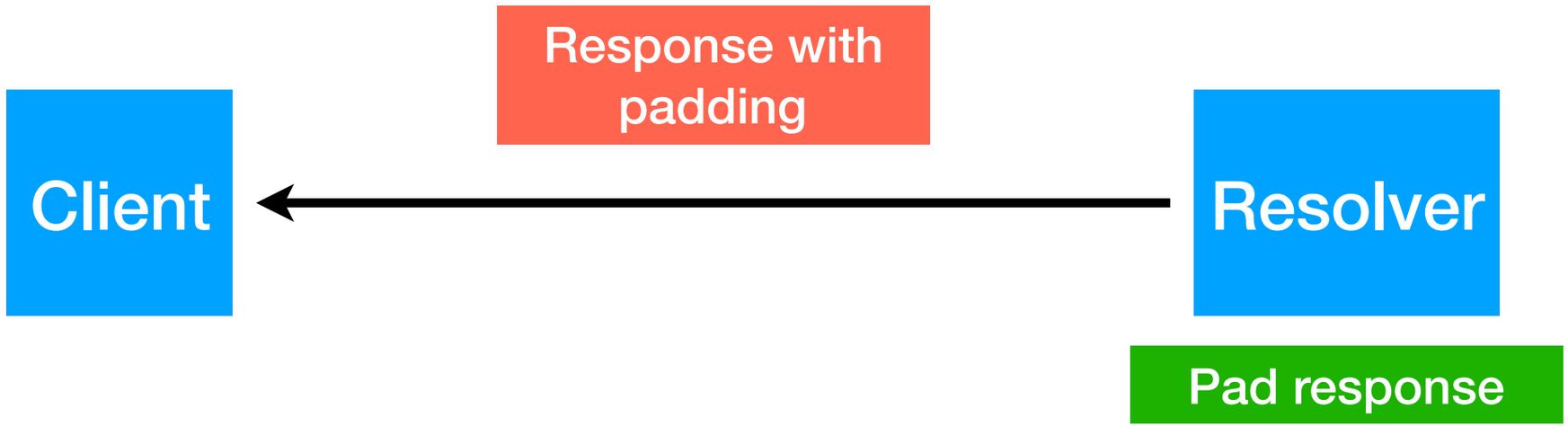
Padding of DNS queries: We implemented the recommended padding strategy[2] on Cloudflare's DoH client. Pad query to multiples of 128 bytes.



[1]RFC7830
[2]RFC8467

EDNS0 Based Countermeasures

Padding of DNS responses: Cloudflare's resolver pads responses to multiples of 128 bytes. Recommended strategy: Pad to multiples of 468 bytes



Our experiments

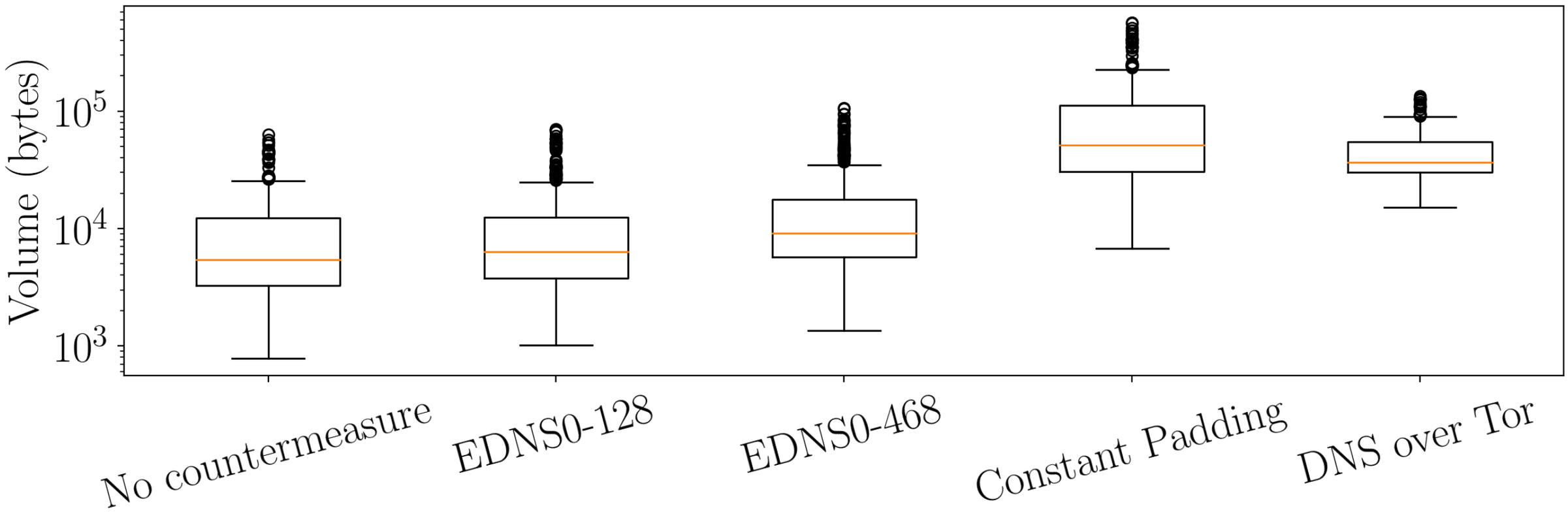
EDNS0-128	Cloudflare’s response padding strategy
EDNS0-468	Recommended response padding strategy
Constant Padding	Keep all TLS record sizes constant
DNS over Tor	Cloudflare’s DNS over Tor service

Results: Classifier performance

Method	Precision	Recall	F1-score
EDNS0-128	0.710 ± 0.005	0.700 ± 0.004	0.691 ± 0.004
EDNS0-468	0.452 ± 0.007	0.448 ± 0.006	0.430 ± 0.007
Constant Padding	0.070 ± 0.003	0.080 ± 0.002	0.066 ± 0.002
DNS over Tor	0.035 ± 0.004	0.037 ± 0.003	0.033 ± 0.003

EDNS0 based measures do not eliminate traffic analysis attacks

Results: Overhead



Sent + received bytes (from TLS records)

Anonymous communication as a defense

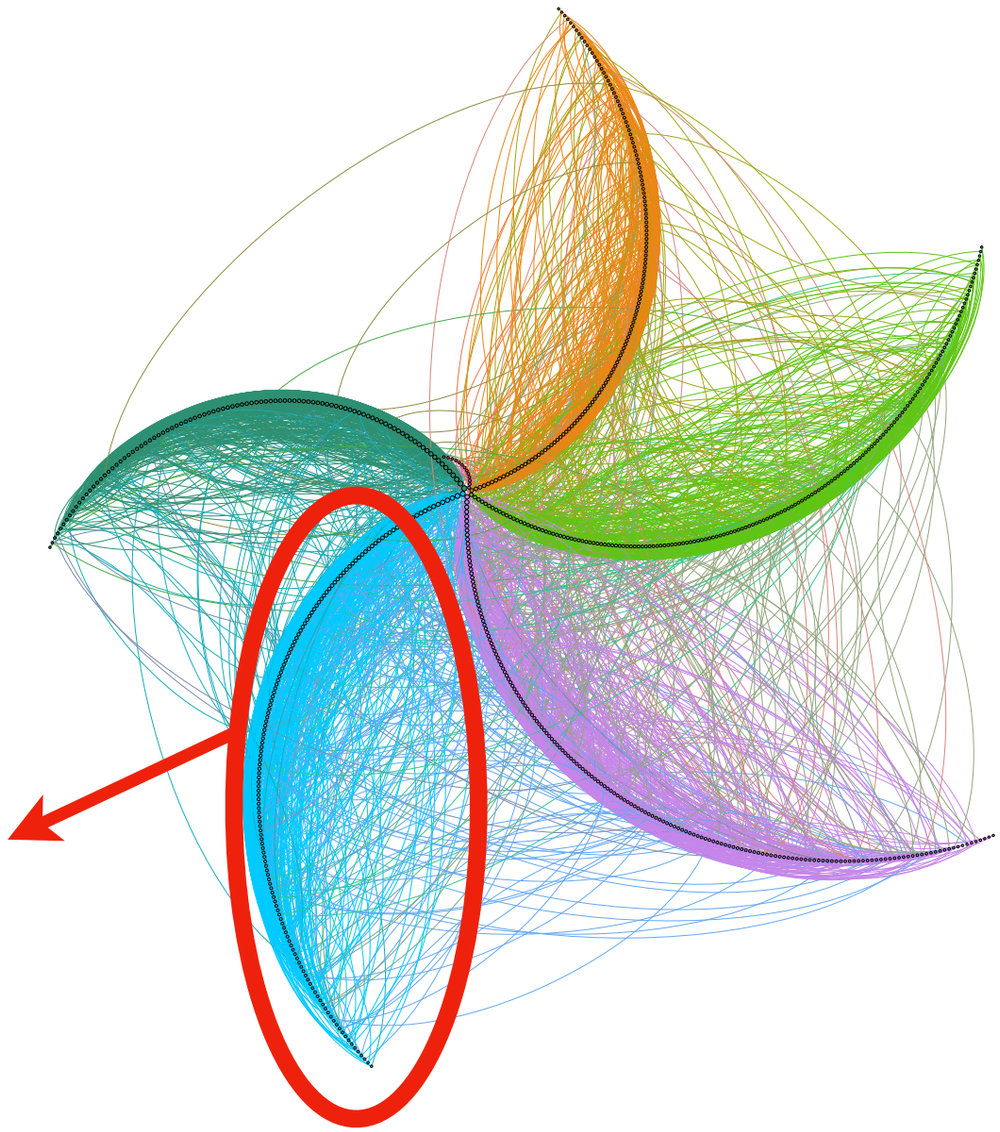
Fixed cell sizes

- 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

Paper preprint: *Encrypted DNS --> Privacy? A Traffic Analysis Perspective* **https://arxiv.org/abs/1906.09682**

Blog post: *Does DoH imply Privacy?* **https://bit.ly/2XXC07t**

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

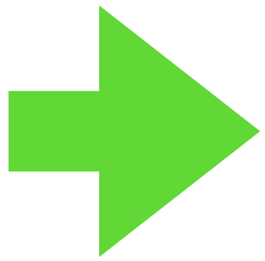
TLS 1.3
Encrypted SNI

```
Length: 17
▼ Server Name Indication extension
    Server Name list length: 15
    Server Name Type: host_name (0)
    Server Name length: 12
    Server Name: sa.bbc.co.uk
```

Feature extraction

TLS record sizes

pcap file



24 -58 63 110 -92 -86 -55

Single record sizes

Burst sizes

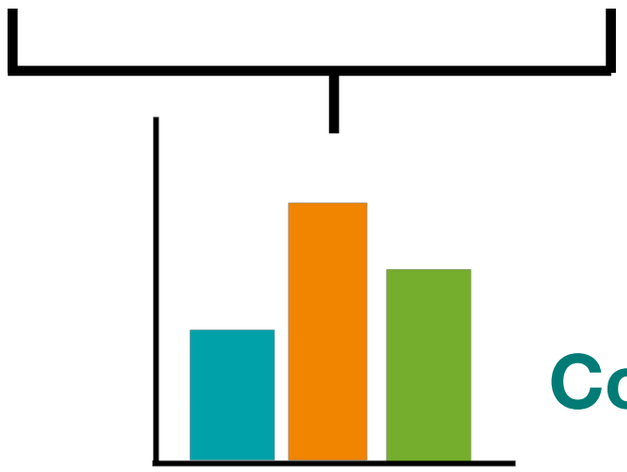
24 -58 173 -233

Uni-grams: (24), (-58)….

Bi-grams: (24, -58), (-58, 63)…

Uni-grams: (24), (-58)…

Bi-grams: (24, -58), (-58, 173)…



Counts