# Avoiding Censorship

Marc Gähwiler

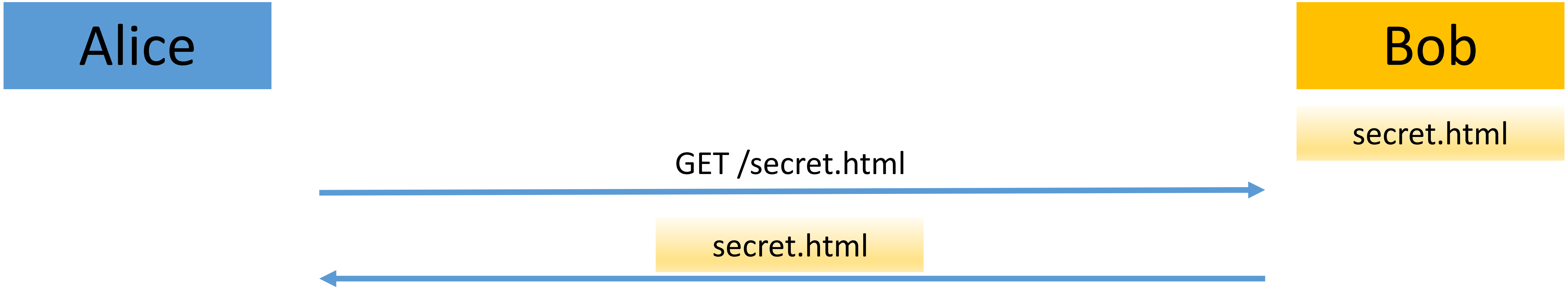
# Motivation

Free & Anonymous Communication

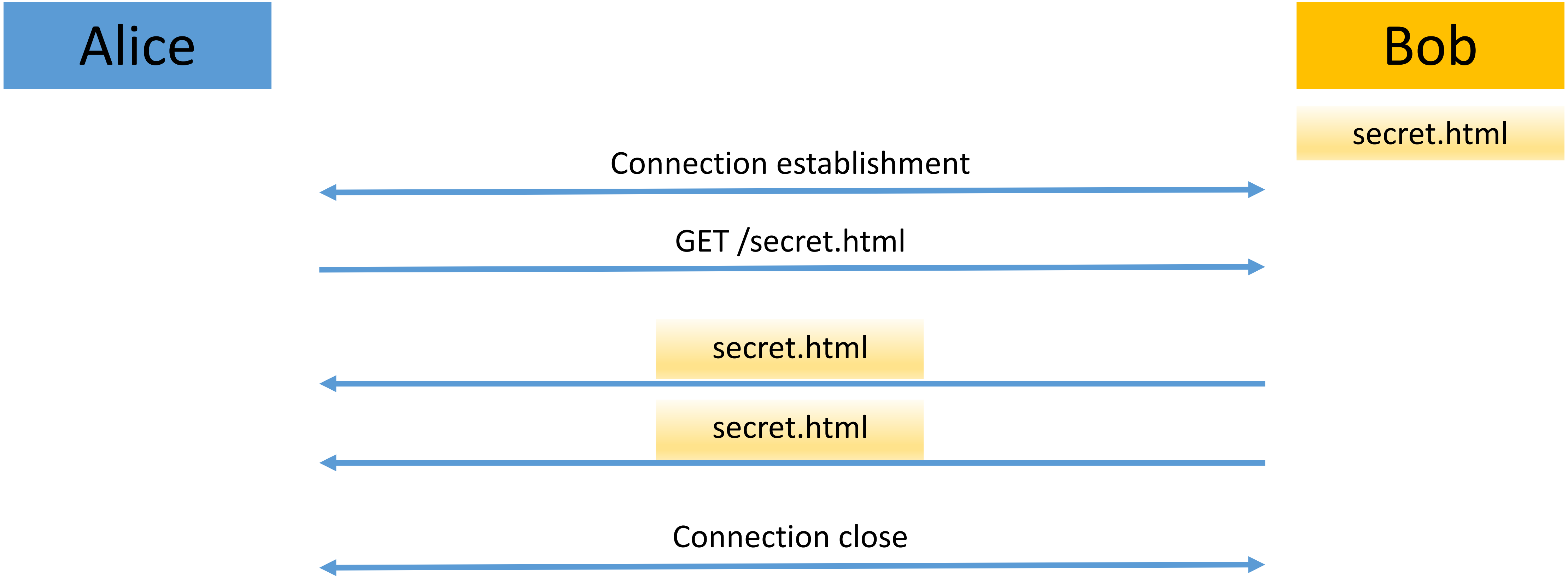
# Background: HTTP / TCP

Alice
Bob
secret.html

# Background: HTTP / TCP



# Background: HTTP / TCP



Alice

Bob

secret.html

Connection establishment

GET /secret.html

secret.html

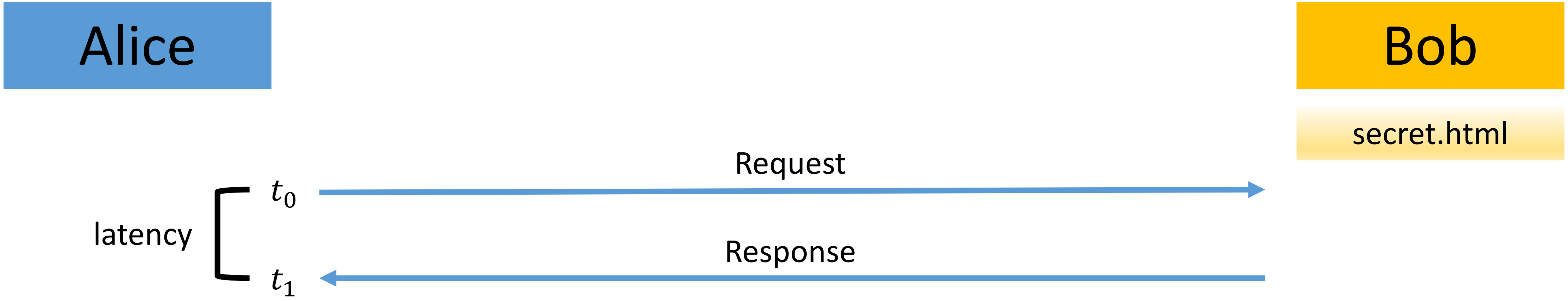
secret.html

Connection close

↔ Multiple TCP packets

→ Single TCP packet (not shown: every packet has to be acknowledged)

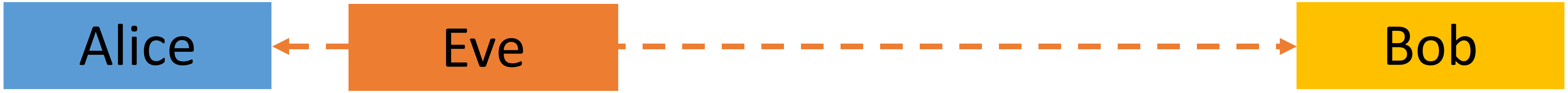
# Background: HTTP / TCP



# Background: TOR The Onion Router & JAP Java Anon Proxy

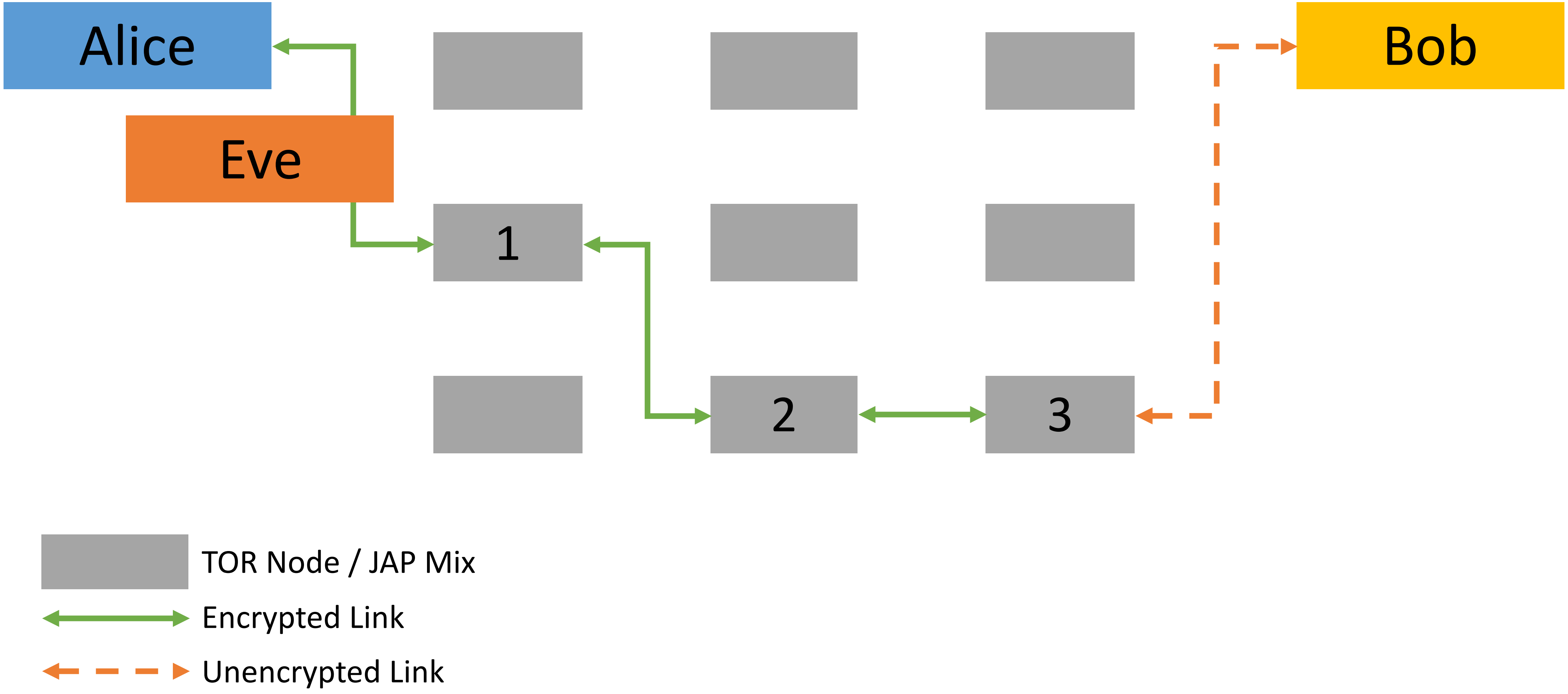
- Anonymous communication
- Hide receiver & content from observer

# Background: TOR The Onion Router & JAP Java Anon Proxy



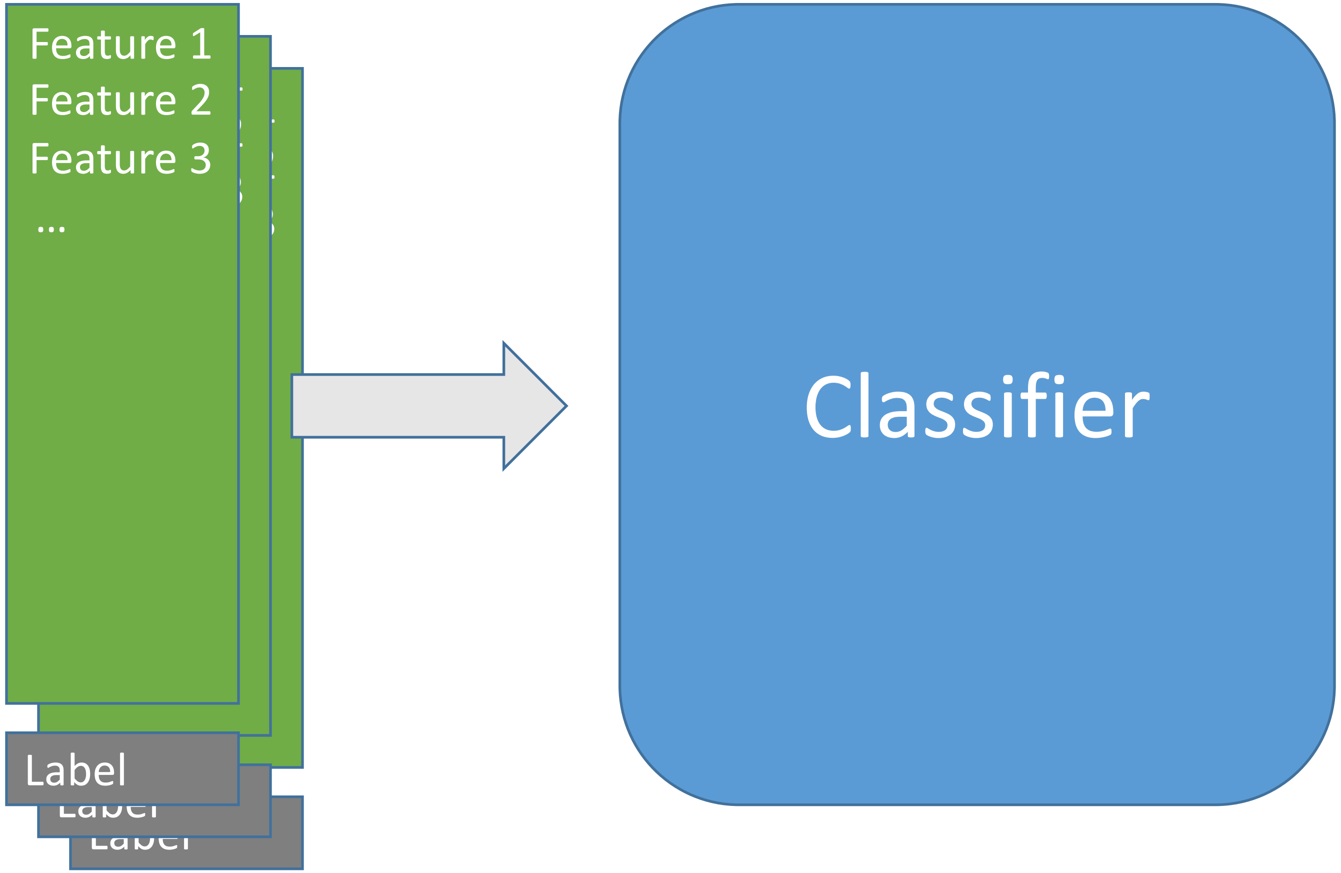
Unencrypted Link

# Background: TOR The Onion Router & JAP Java Anon Proxy



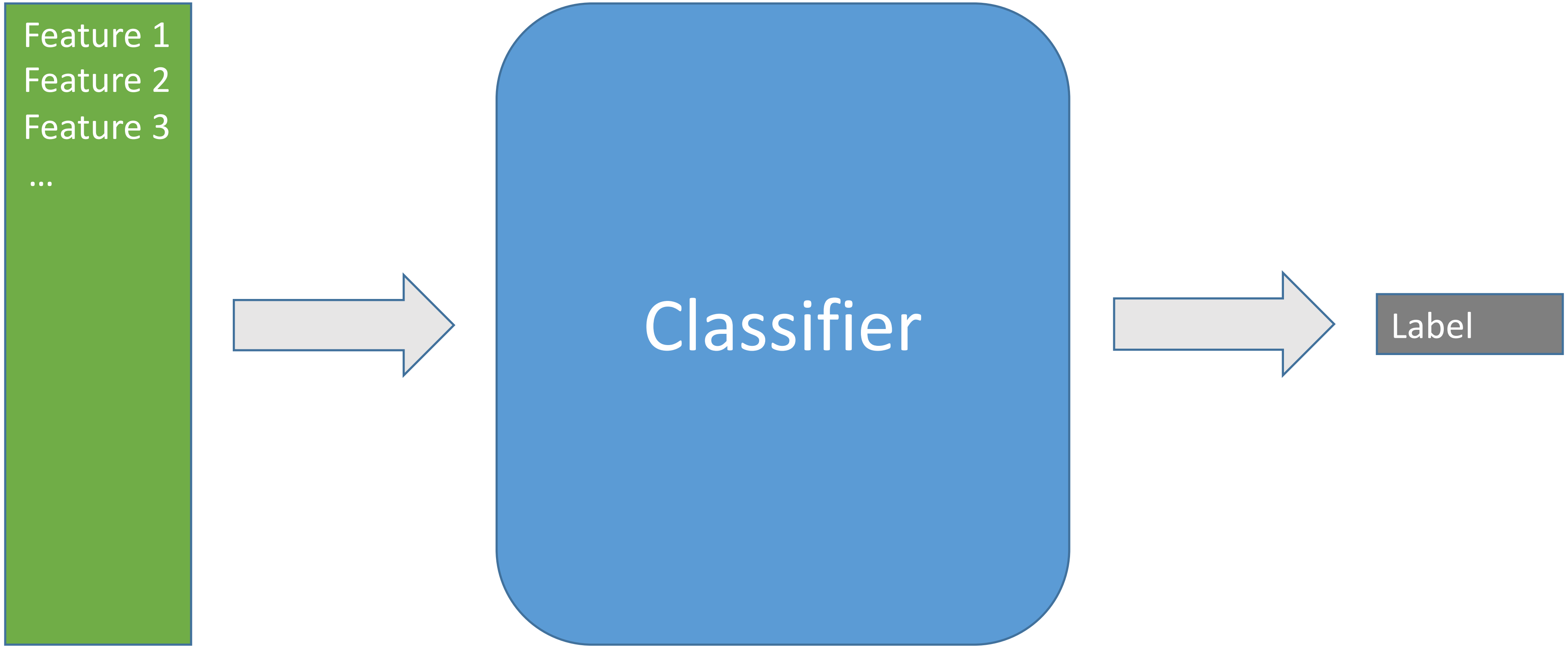
# Background: Machine Learning

Training



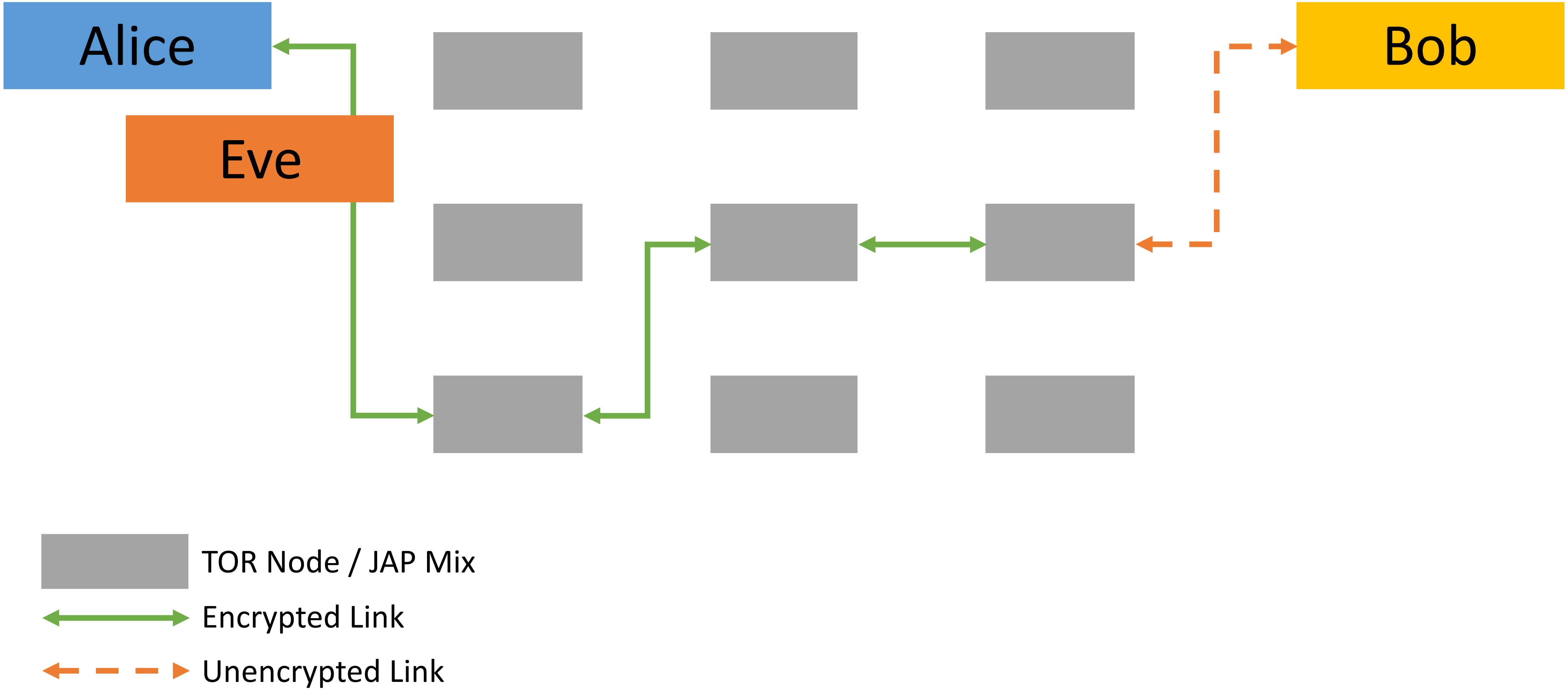
# Background: Machine Learning

## Prediction

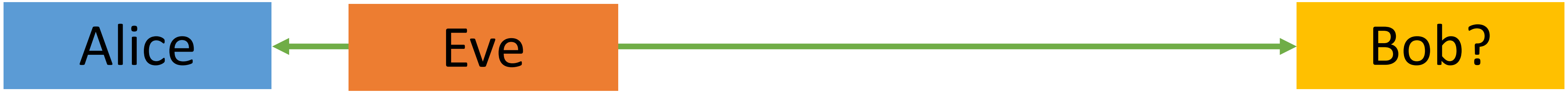


# Website Fingerprinting in Onion Routing Based Anonymization Networks

# Website Fingerprinting: Idea



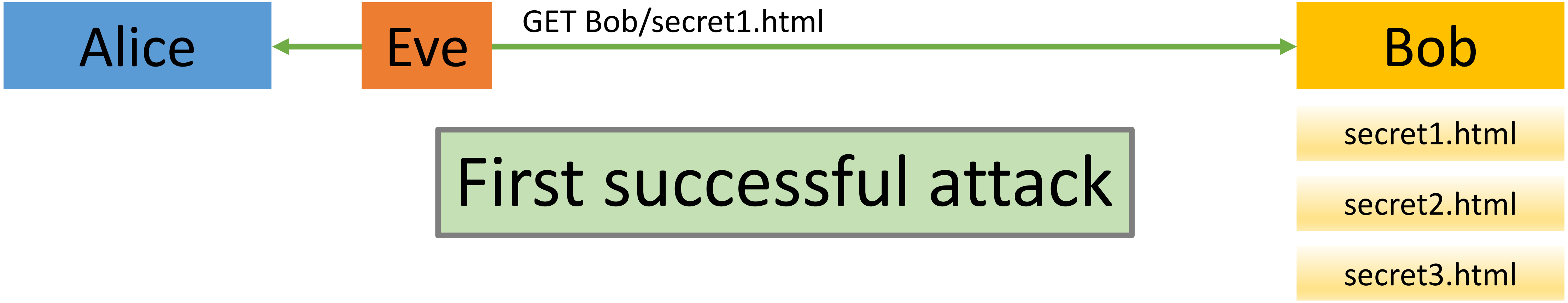
# Website Fingerprinting: Idea



Eve

- Local eavesdropper
- Can analyse traffic
  - Volume of transferred data
  - Packet timings / sizes
  - …
- Goal: Recognize requested web-page

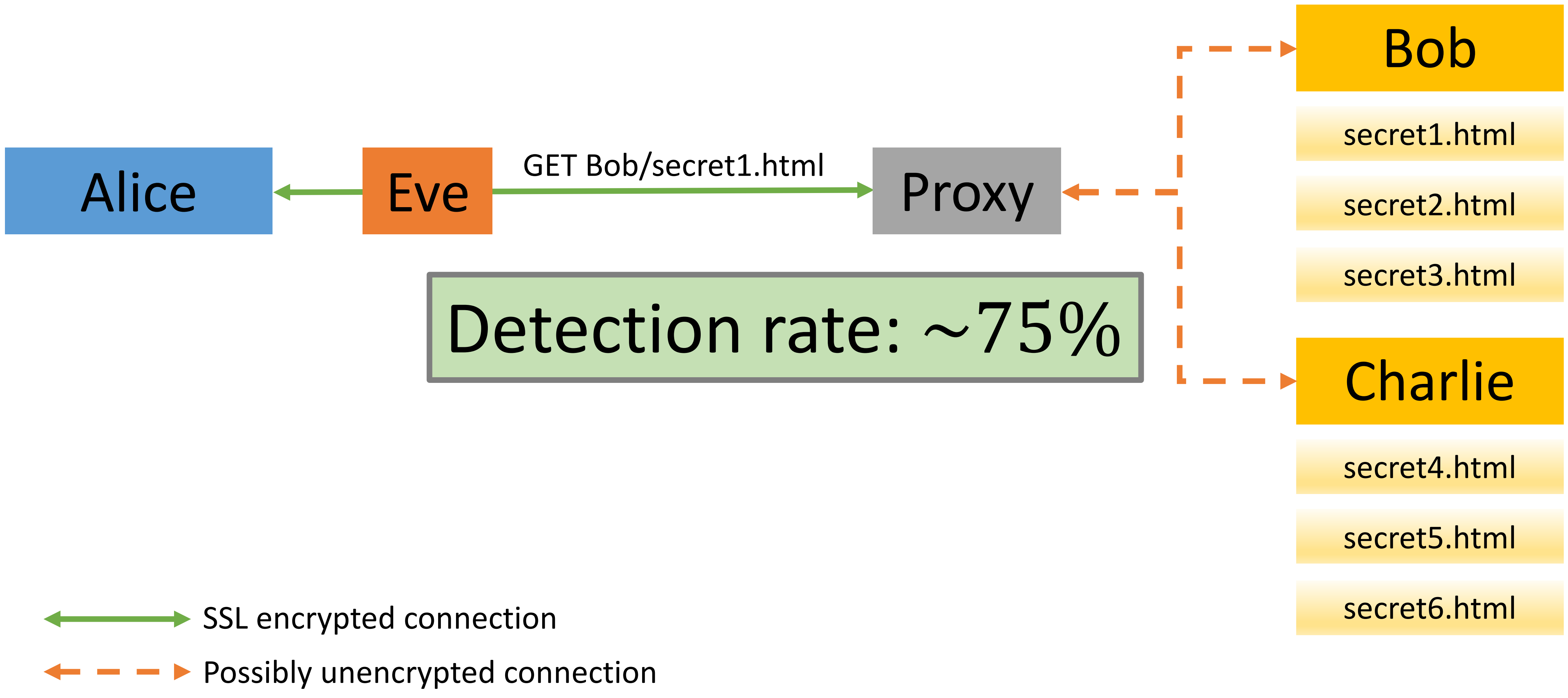
# Website Fingerprinting: Earlier Work 1



First successful attack

⟷ SSL encrypted connection

# Website Fingerprinting: Earlier Work 2



Alice

Eve

GET Bob/secret1.html

Proxy

Bob

secret1.html

secret2.html

secret3.html

Charlie

secret4.html

secret5.html

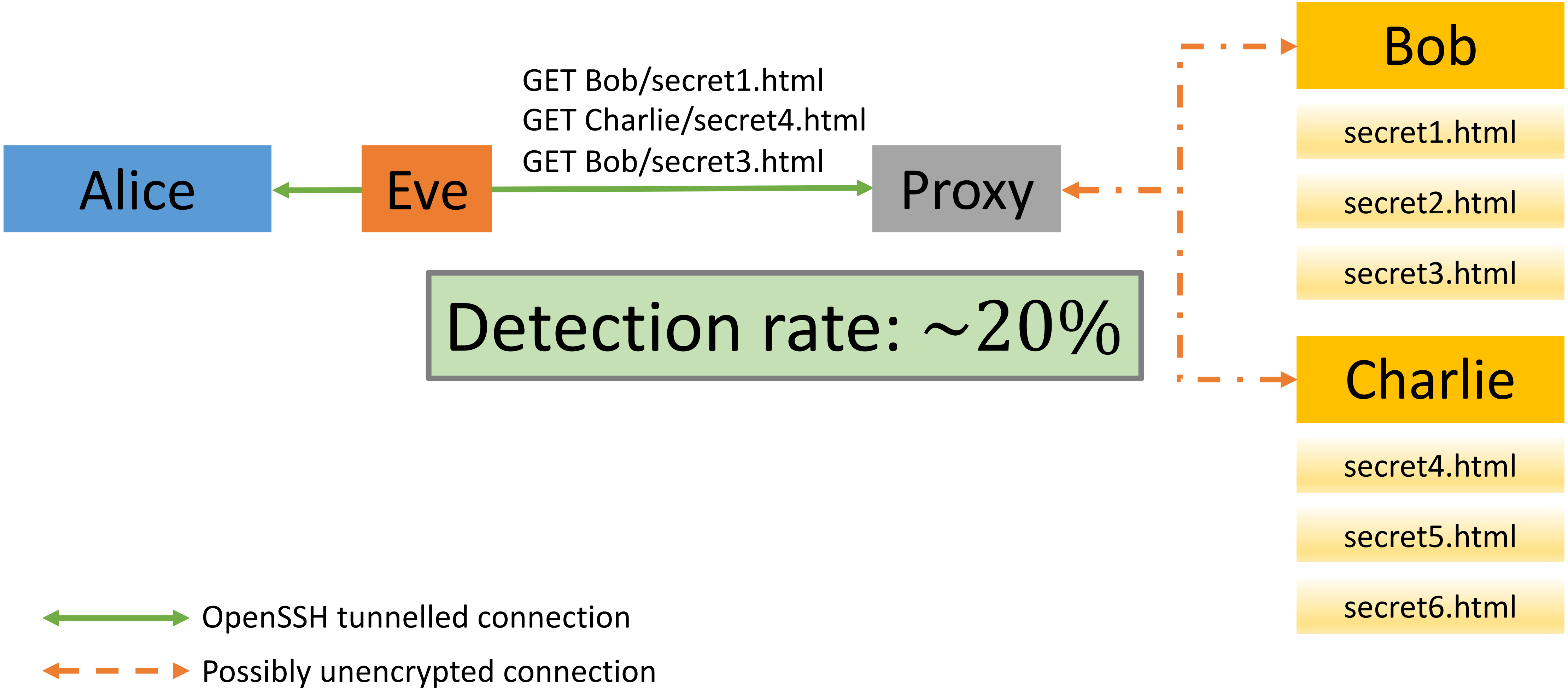
secret6.html

Detection rate: ~75%

SSL encrypted connection

Possibly unencrypted connection

# Website Fingerprinting: Earlier Work 3



Alice

Eve

GET Bob/secret1.html
GET Charlie/secret4.html
GET Bob/secret3.html

Proxy

Bob

secret1.html

secret2.html

secret3.html

Charlie

secret4.html

secret5.html

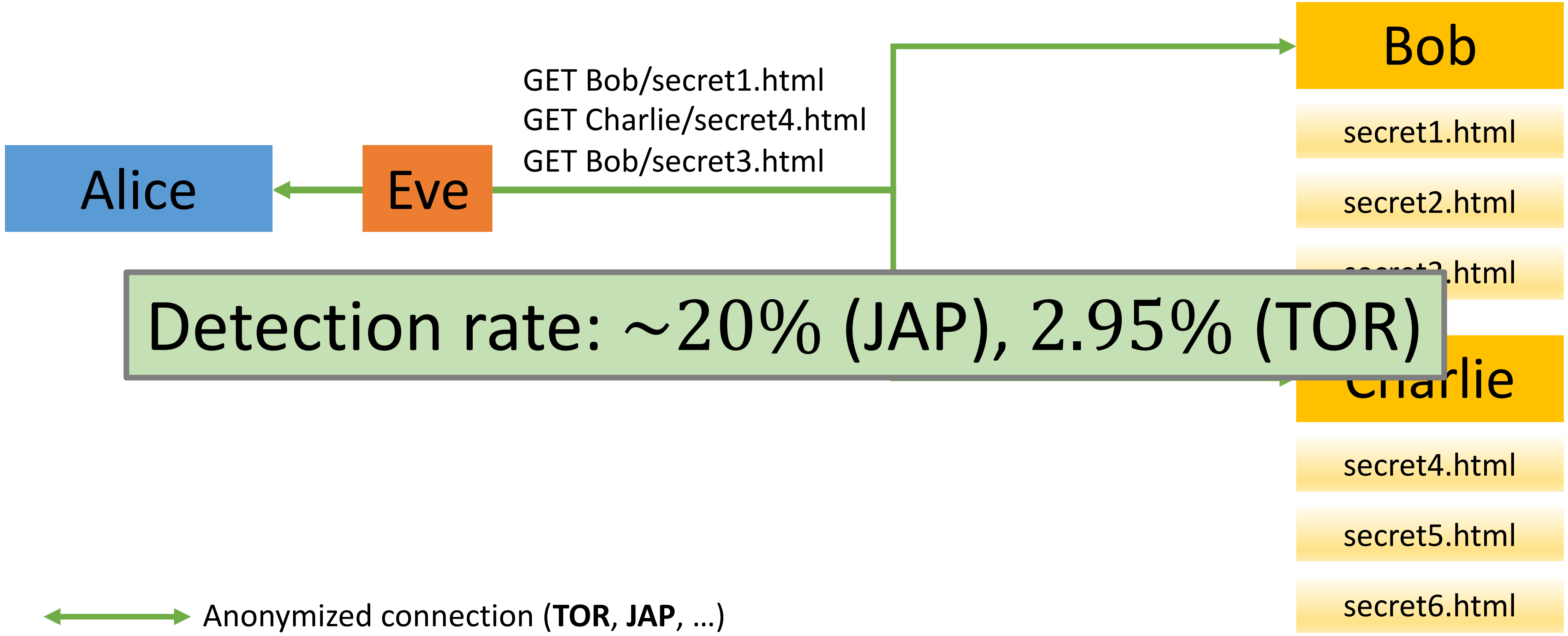
secret6.html

Detection rate: ~20%

OpenSSH tunnelled connection

Possibly unencrypted connection

# Website Fingerprinting: Earlier Work 4



Alice

Eve

GET Bob/secret1.html
GET Charlie/secret4.html
GET Bob/secret3.html

Bob

secret1.html

secret2.html

secret3.html

Charlie

secret4.html

secret5.html

secret6.html

Detection rate: ~20% (JAP), 2.95% (TOR)

Anonymized connection (**TOR**, **JAP**, …)

# Website Fingerprinting: New Approach – Features

- Feature selection is crucial
- Previous work: Packet size & packet direction
- This paper: Find important features

# Website Fingerprinting: New Approach – Features

- Without Packets Sized 52
- Size Markers
- HTML Markers
- Total Transmitted Bytes
- Number Markers
- Occurring Packet Sizes
- Percentage Incoming Packets
- Number Of Packets

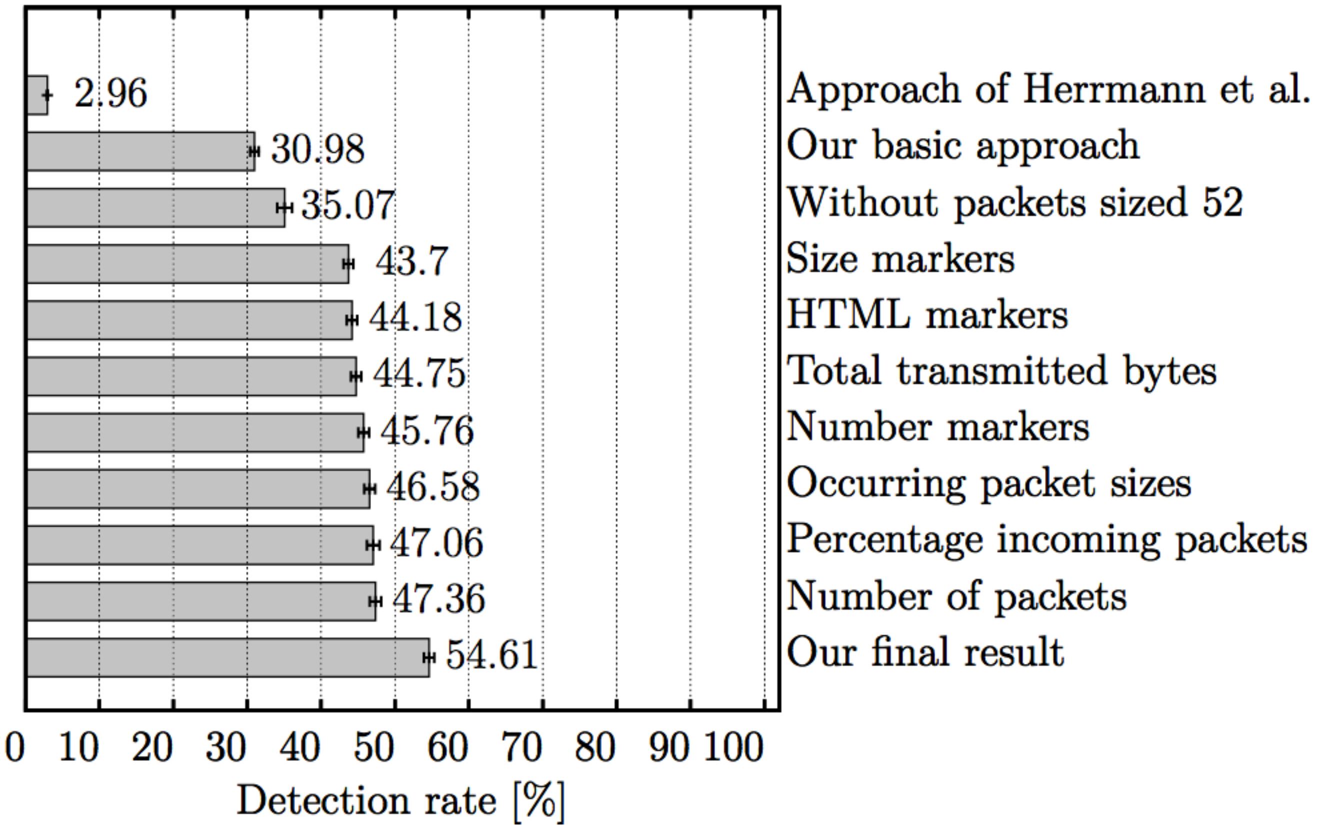
# Website Fingerprinting:
# New Approach – Improved classification

- Support vector machines (SVM)
- Optimized SVM parameters

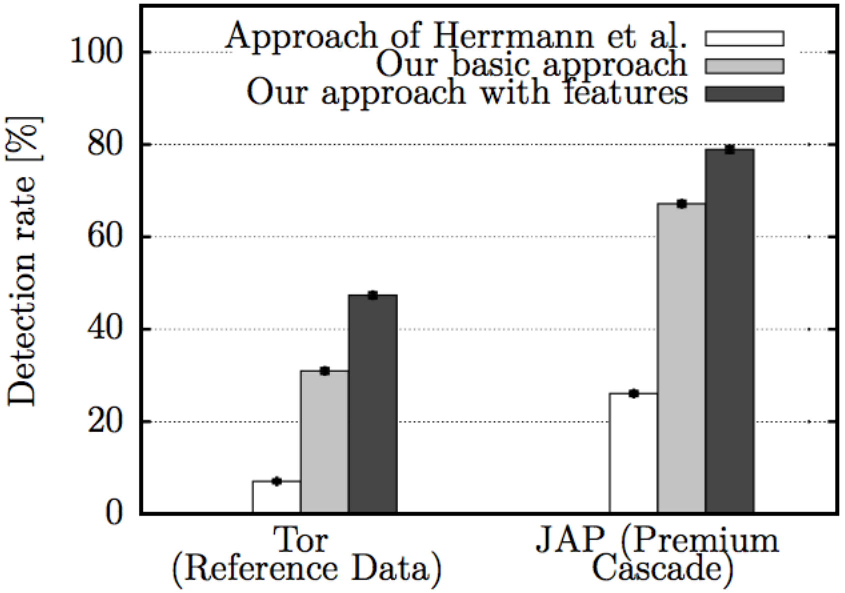
# Website Fingerprinting: Closed-World Results

- 775 different web pages
- Redirect → final page
- Incomplete page → Reload

# Website Fingerprinting: Closed-World Results



# Website Fingerprinting: Closed-World Results



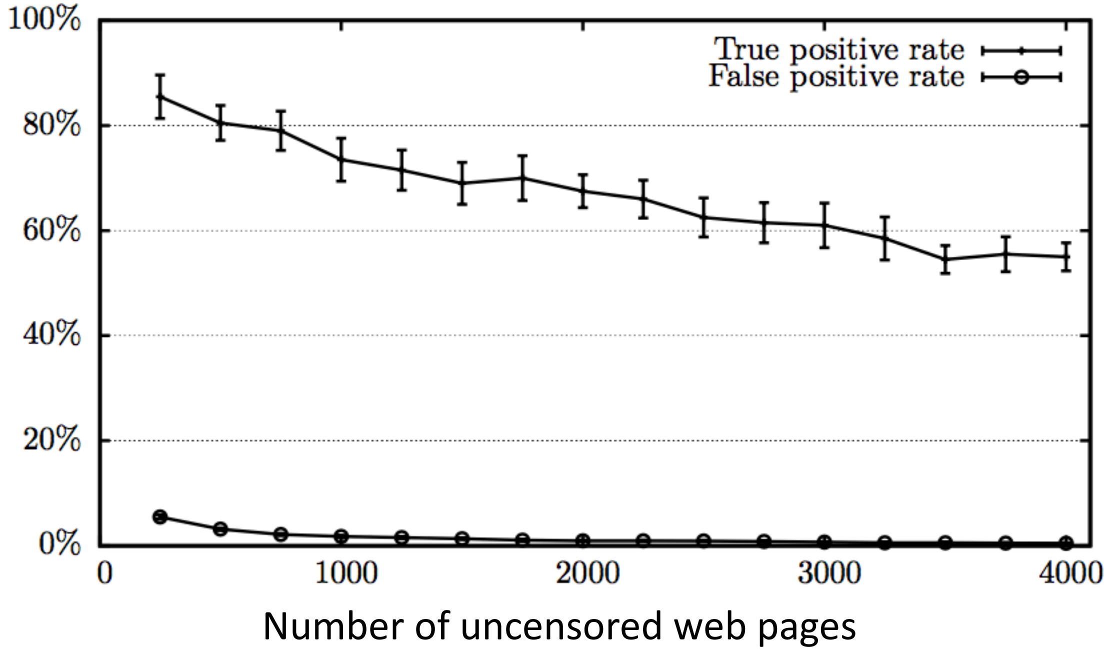
# Website Fingerprinting: Open-World Results

- Censored
  - 3 lists: ”Sexually Explicit”, “Alexa Top Ranked”, “Alexa Random”
  - Training: 5 random URLs out of list (35 instances each)
  - Testing: same 5 URLs (25 instances each)
- Uncensored
  - 1,000,000 most popular pages
  - Training: 4,000 random URLs (1 instance each)
  - Testing: 1,000 random URLs (1 instance each)

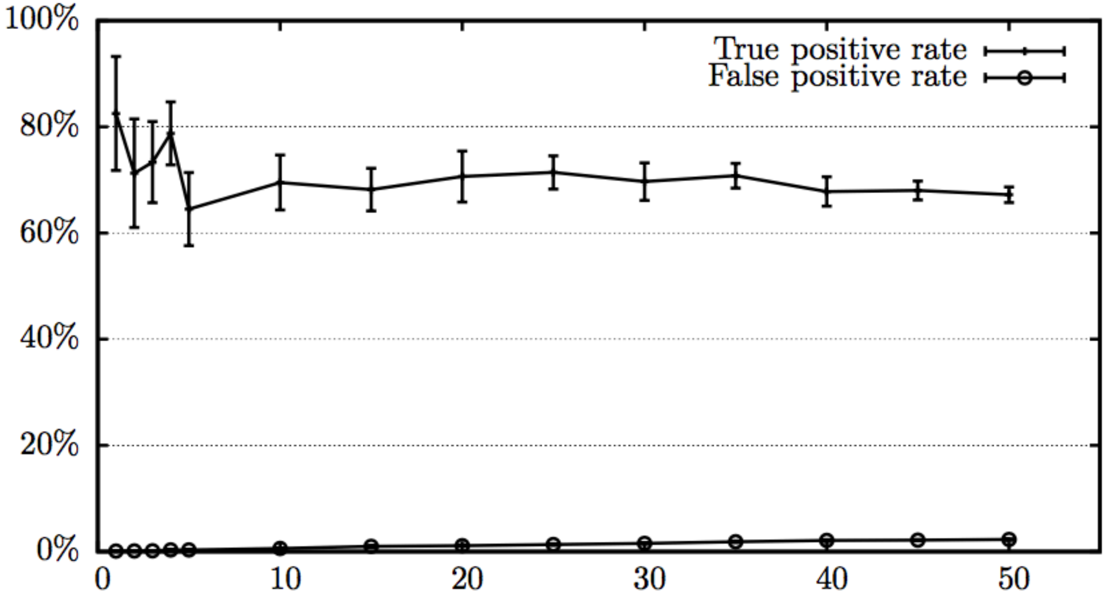
# Website Fingerprinting: Open-World Results

| Page Set | True Positives | False Positives |
|---|---|---|
| Sexually explicit | 56.0% | 0.89% |
| Alexa top ranked | 73.0% | 0.05% |
| Alexa random | 56.5% | 0.23% |

# Website Fingerprinting: Open-World Results

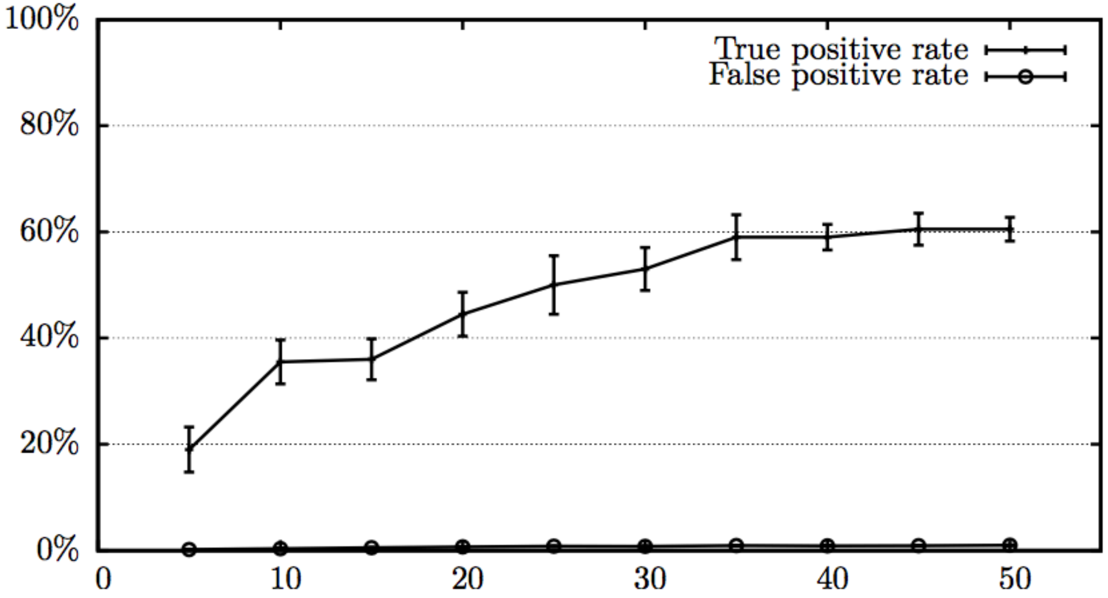


# Website Fingerprinting: Open-World Results



Number of censored web pages

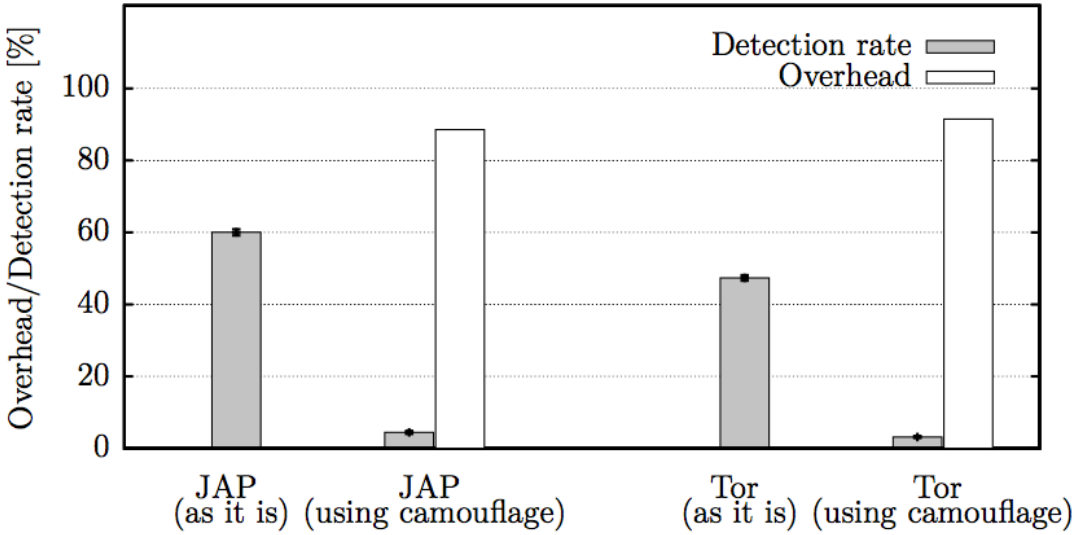
# Website Fingerprinting: Open-World Results



Number of instances of censored web pages

# Website Fingerprinting: Countermeasures

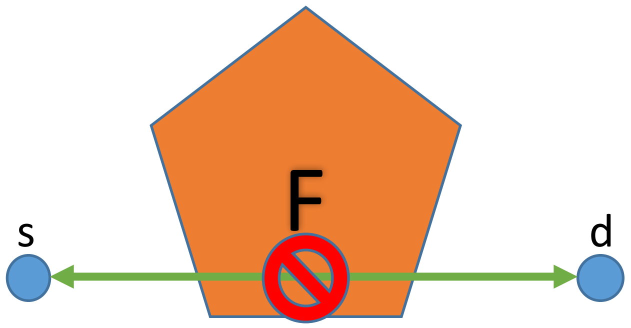
- TOR & JAP use padding
- Proposed countermeasure:
  - Simultaneously load random page



# Alibi Routing

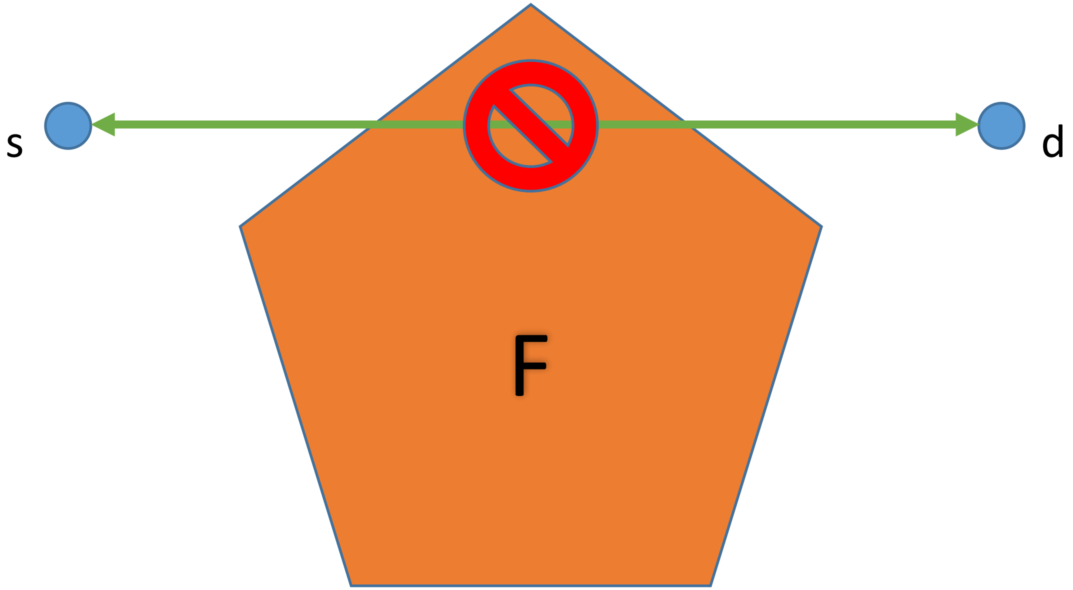
# Alibi Routing: Idea

- Proof of avoidance
- No hardware/policy modifications
- Use
  - GPS coordinates
  - Speed of light

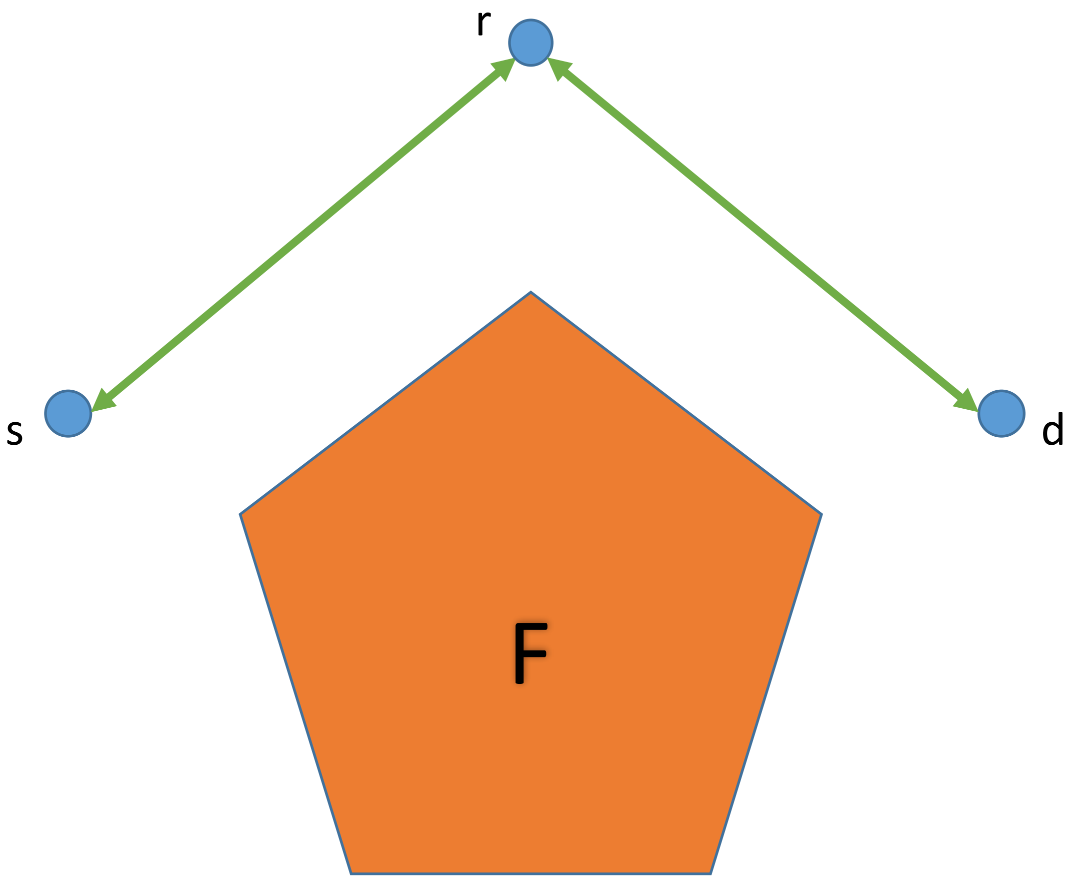


Result: Routing system to avoid geographical regions

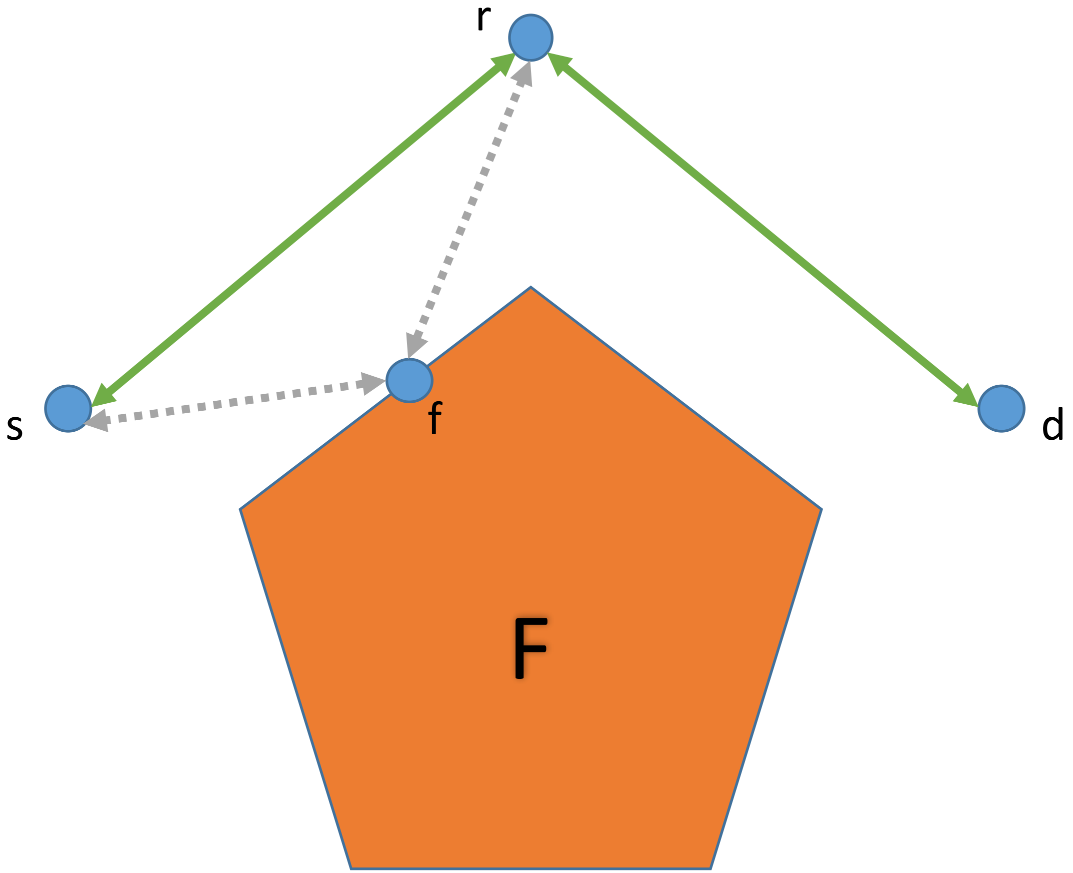
# Alibi Routing: "Proof of avoidance"



# Alibi Routing: “Proof of avoidance”



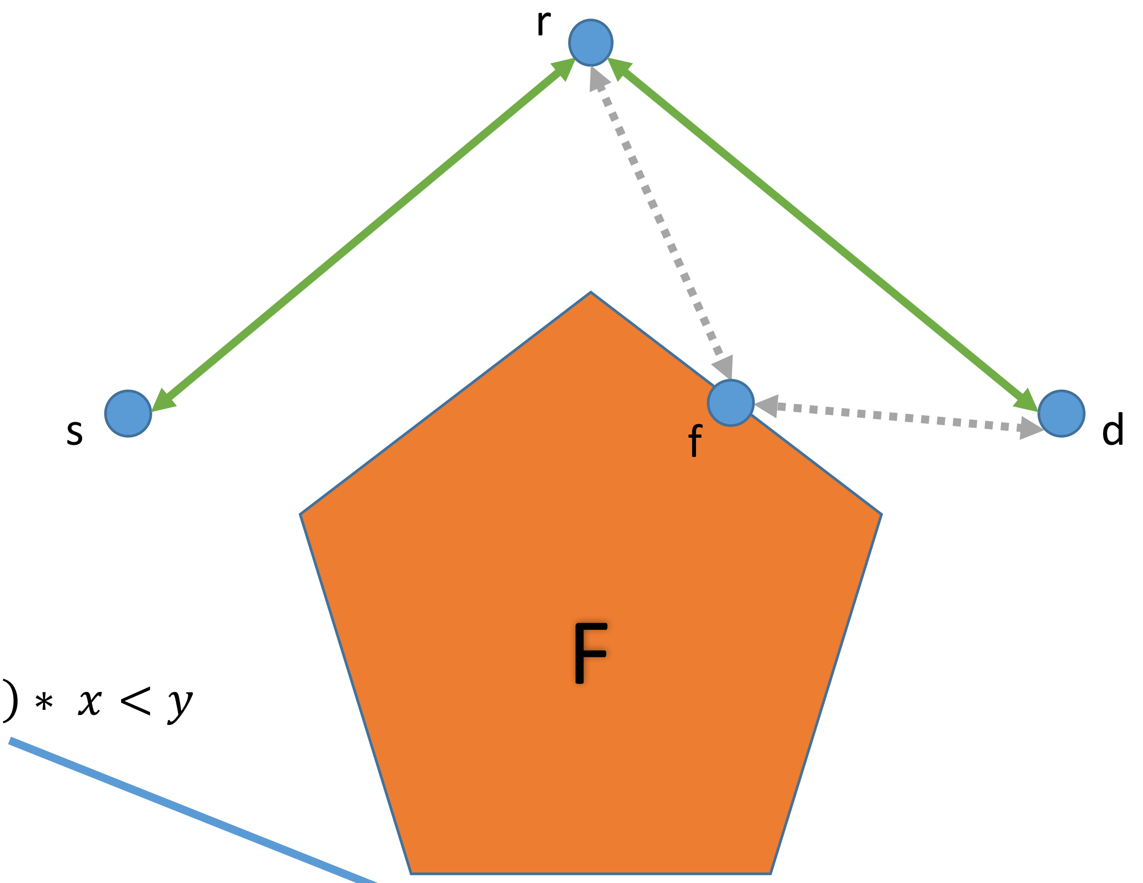
# Alibi Routing: “Proof of avoidance”



$$R(s,r) + R(r,d) \ll \min_{f \in F}\{R(s,f) + R(f,r)\} + R(r,d)$$

$$\rightarrow R(s,r) \ll \min_{f \in F}\{R(s,f) + R(f,r)\}$$

# Alibi Routing: “Proof of avoidance”



$x \ll y$:
For some $\delta \geq 0$: $(1+\delta) * \ x < y$

$$R(s,r) + R(r,d) \ll R(s,r) + \min_{f \in F}\{R(r,f) + R(f,d)\}$$

# Alibi Routing: Protocol

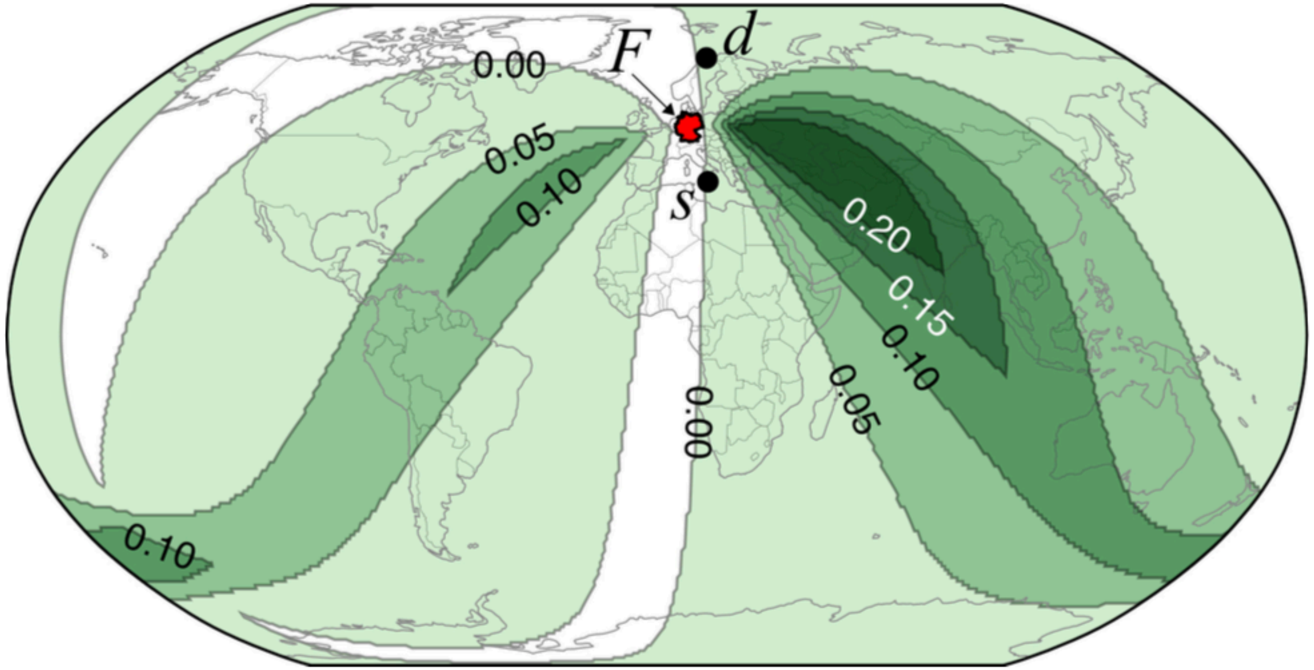
Assumptions/facts

- Peers outside $F$ are trustworthy
- No lies about lower latency
- Speed of light

# Alibi Routing: Protocol

Query: $\langle s, d, F, T \rangle$

- $s$: source
- $d$: destination
- $F$: forbidden regions
- $T$: target regions

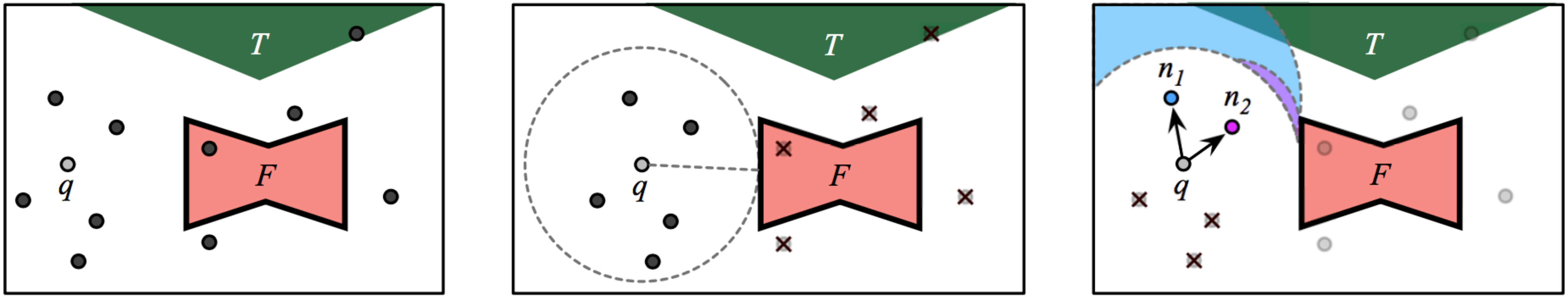


Target region

- Node $g$ is in $T$ if
  $(1 + \delta) \cdot D(s, g) < \min_{f \in F}\{D(s, f) + D(f, g)\}$ and
  $(1 + \delta) \cdot D(g, d) < \min_{f \in F}\{D(g, f) + D(f, d)\}$

# Alibi Routing: Protocol

- Node maintains sets of
  - Known active peers
  - Neighbours used to process queries
- Task: Determine next-hop neighbour & forward query

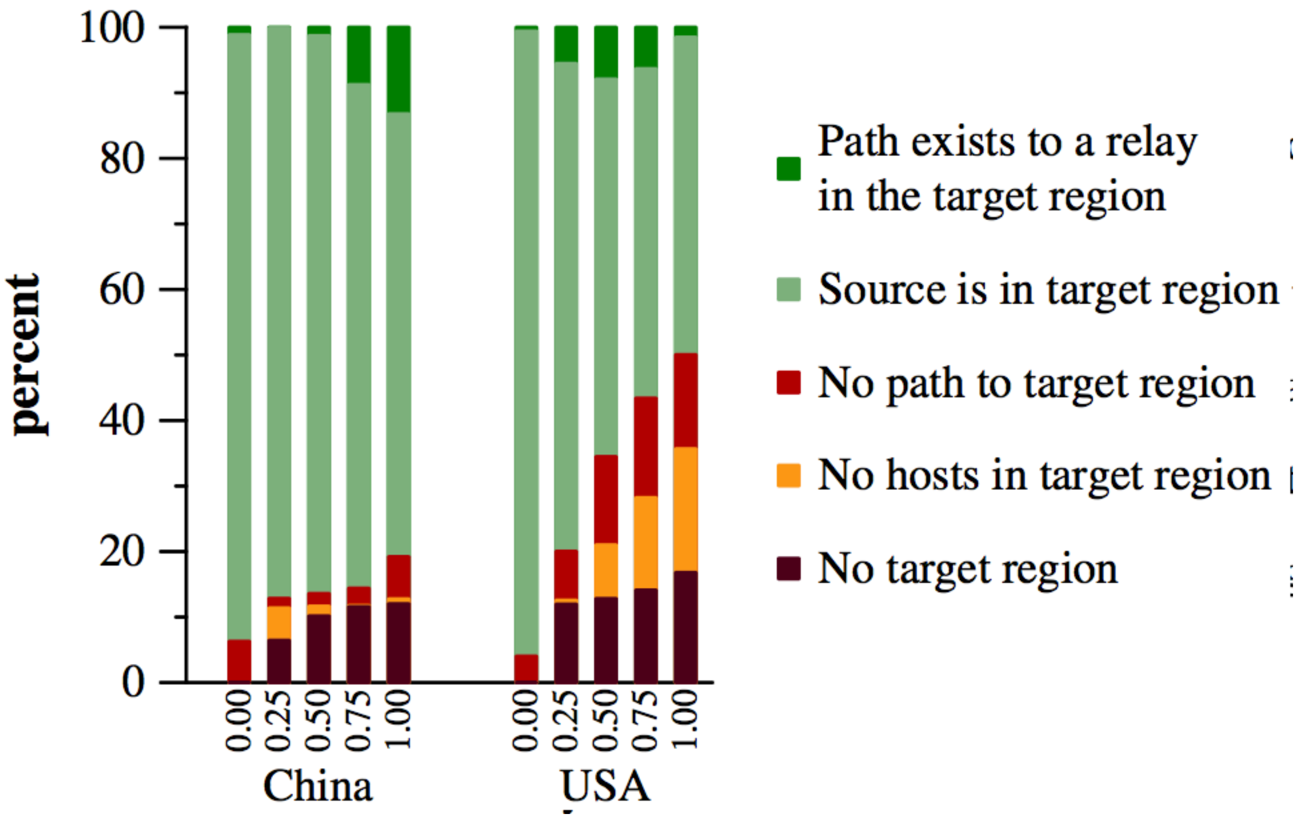


# Alibi Routing: Security

- Safety
- Progress
- Non-attacks

# Alibi Routing: Evaluation – Feasibility

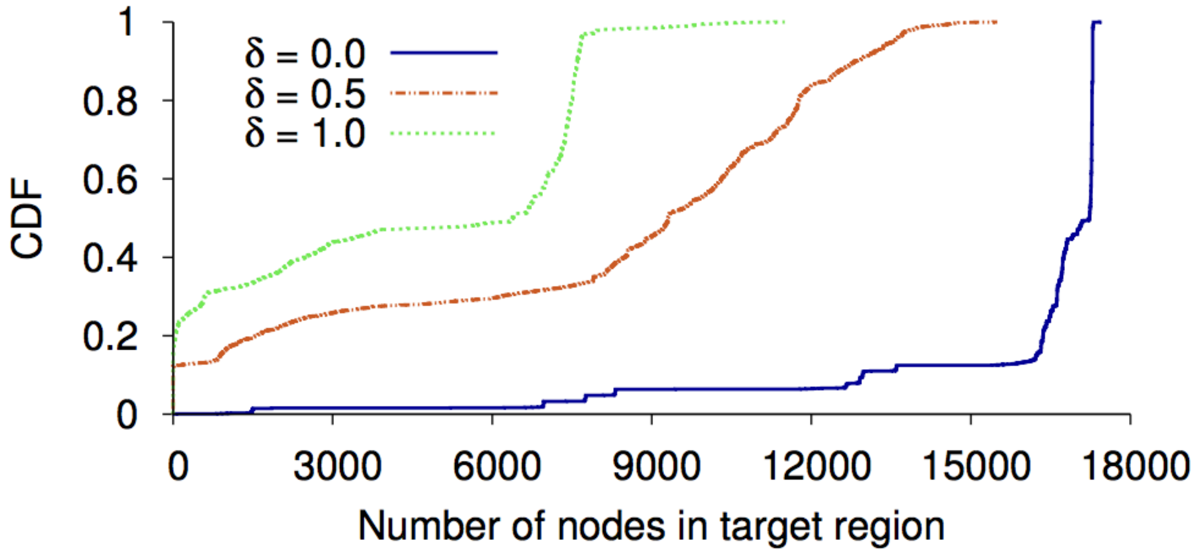
Can source reach destination?



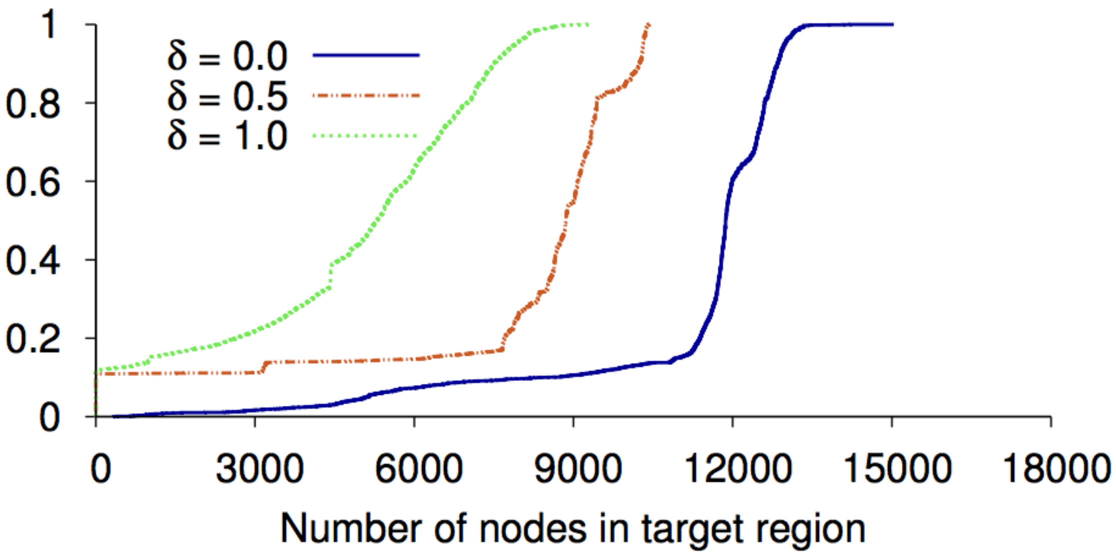
# Alibi Routing: Evaluation – Feasibility

## Target region size

USA is forbidden region



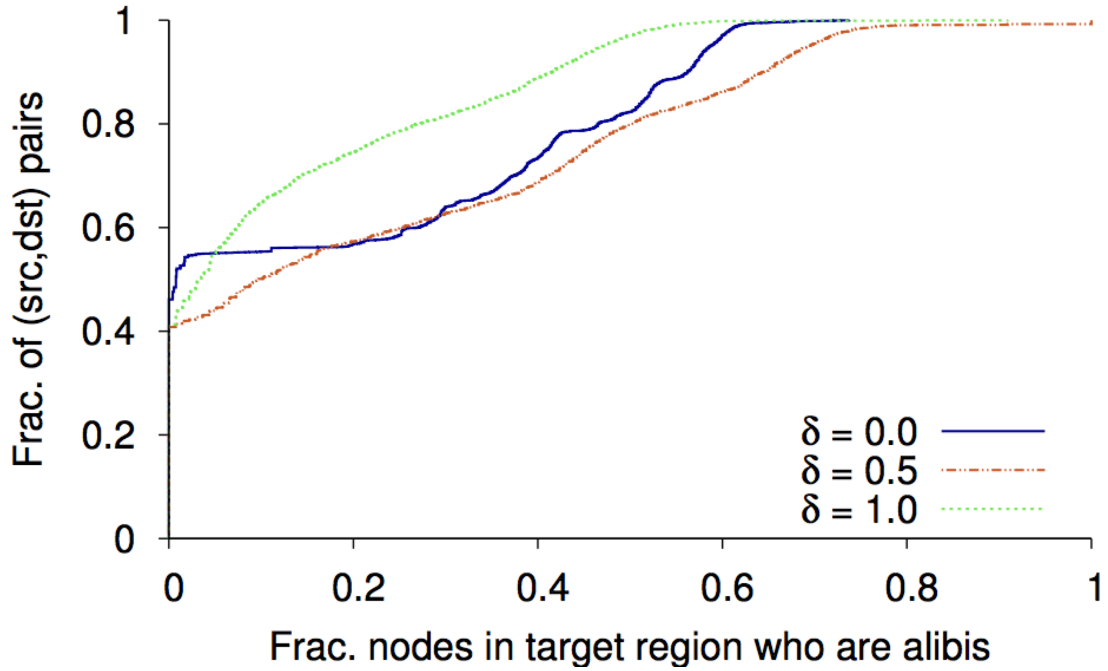
China is forbidden region



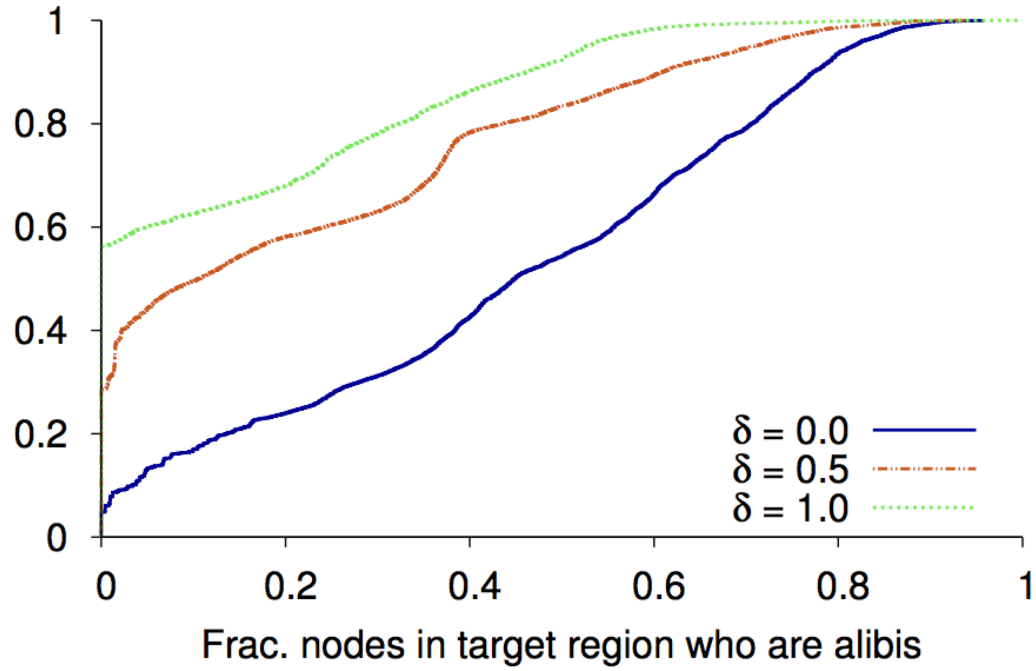
# Alibi Routing: Evaluation – Feasibility

## Viable alibis in target region

USA is forbidden region



China is forbidden region



# Alibi Routing: Evaluation – Performance

## Success & overhead

Success rate

| | Number of nodes | |
|---|---|---|
| $\delta$ | 10,000 | 20,000 |
| 0 | 99.5% | 100% |
| 0.5 | 84.12% | 93.60% |
| 1.0 | 84.12% | 93.28% |

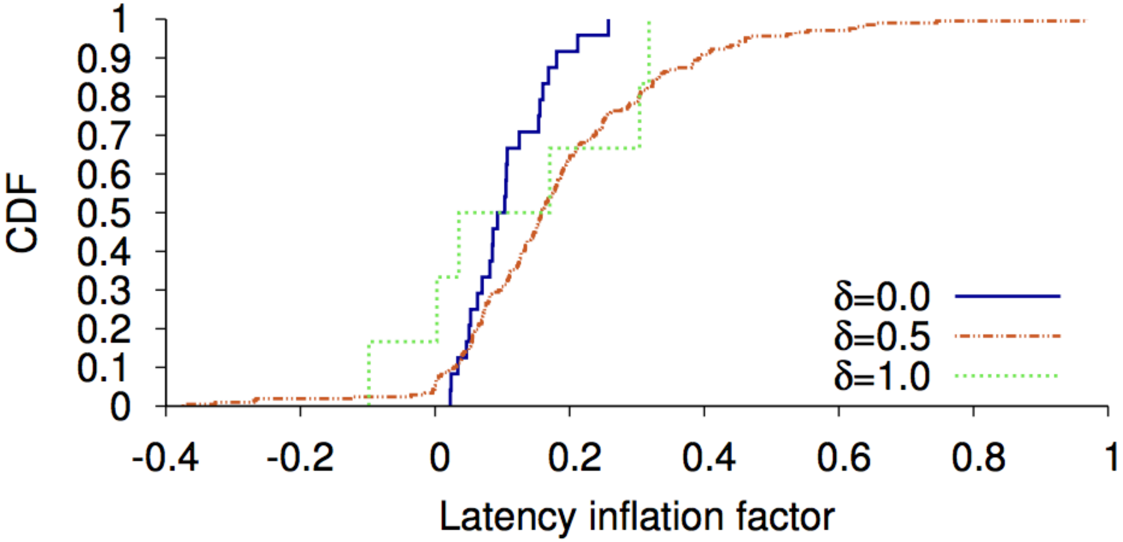
Average number of nodes contacted

| | Number of nodes | |
|---|---|---|
| $\delta$ | 10,000 | 20,000 |
| 0 | 7.11 | 4.68 |
| 0.5 | 44.40 | 37.14 |
| 1.0 | 38.76 | 35.58 |

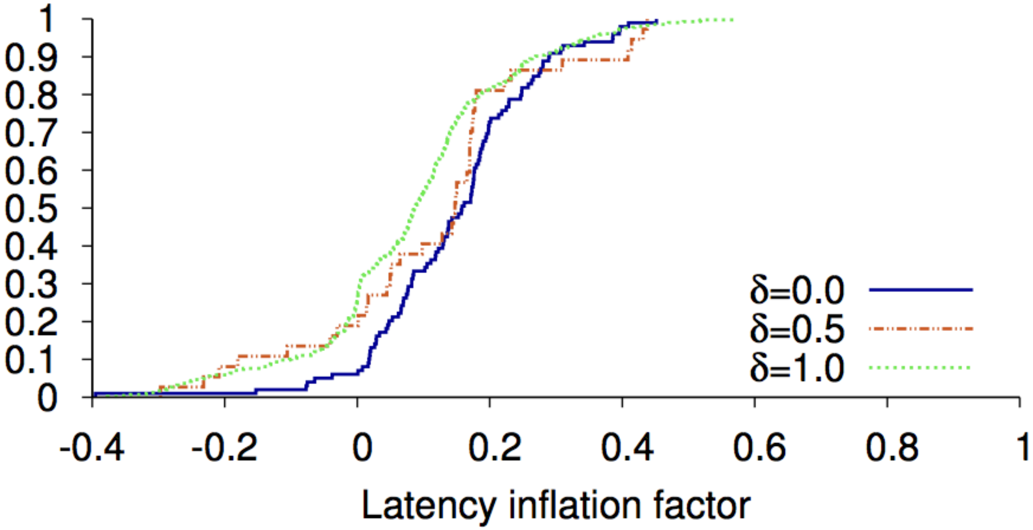
# Alibi Routing: Evaluation – Performance

## Latency inflation

USA is forbidden region



China is forbidden region



# Summary

Website fingerprinting

- Weak anonymity of TOR & JAP
- Countermeasure

Alibi Routing

- Provable avoidance routing scheme