Website Fingerprinting

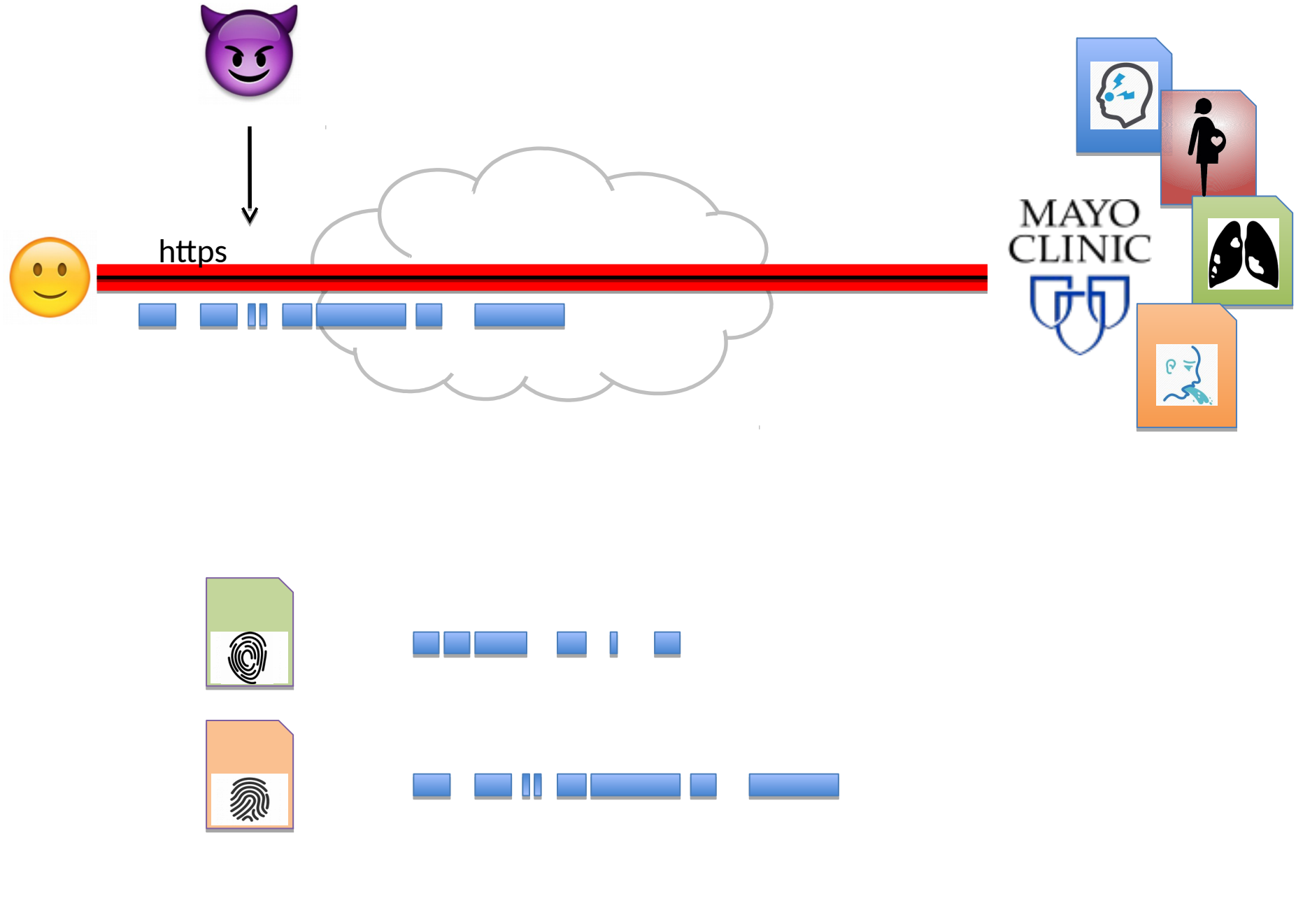
Claudia Diaz
KU Leuven – COSIC

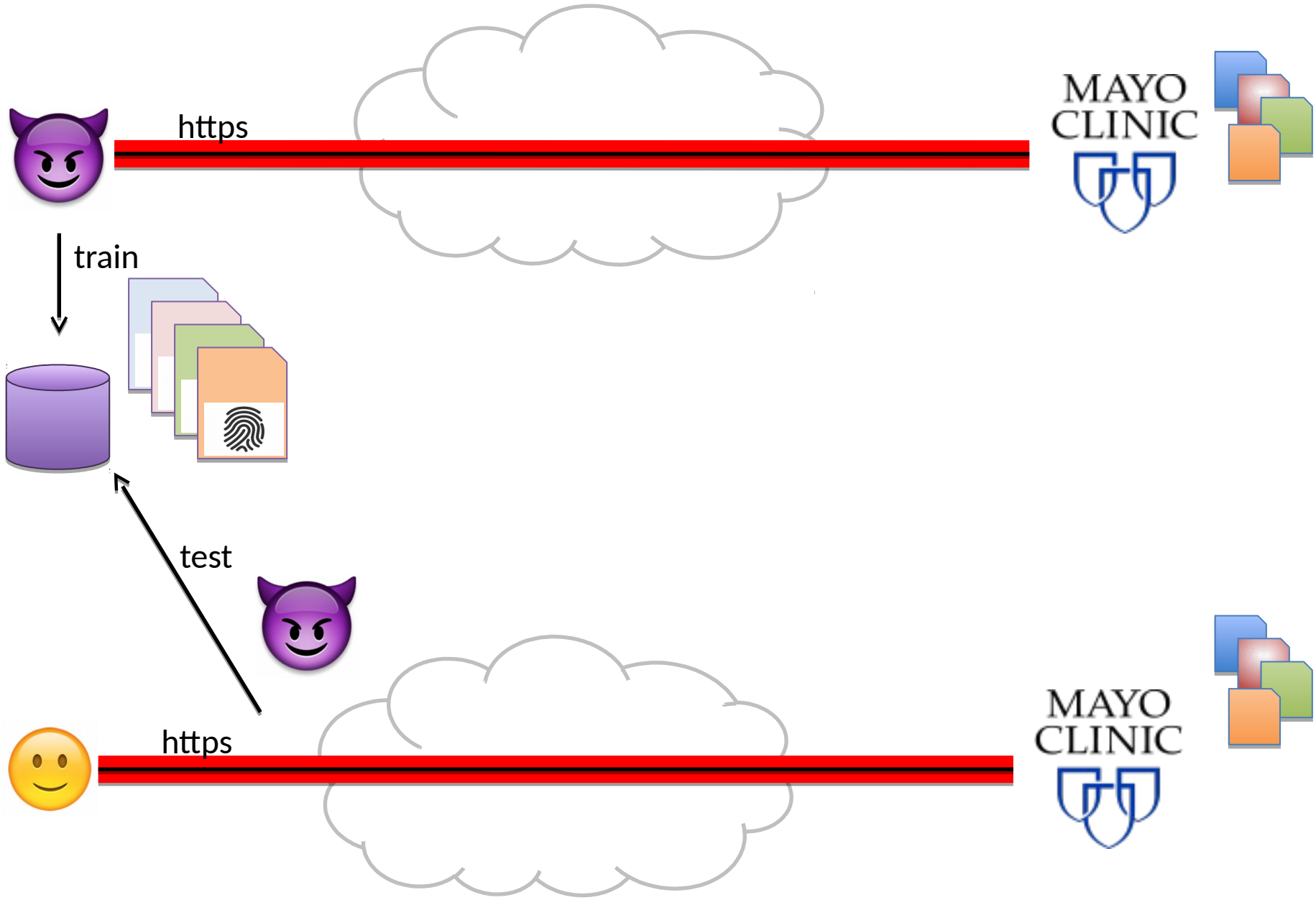
(With thanks to Marc Juarez and Bekah Overdorf)

Summer School on real-world crypto and privacy
June 2017

Outline

- Website Fingerprinting for https sites
- Website Fingerprinting for Tor
- From the lab to reality: reviewing assumptions
- Fingerprintability of hidden services


https
MAYO
CLINIC


https
MAYO CLINIC
train
test
https
MAYO CLINIC

Side channel leaks in web applications (Chen et al, 2010)

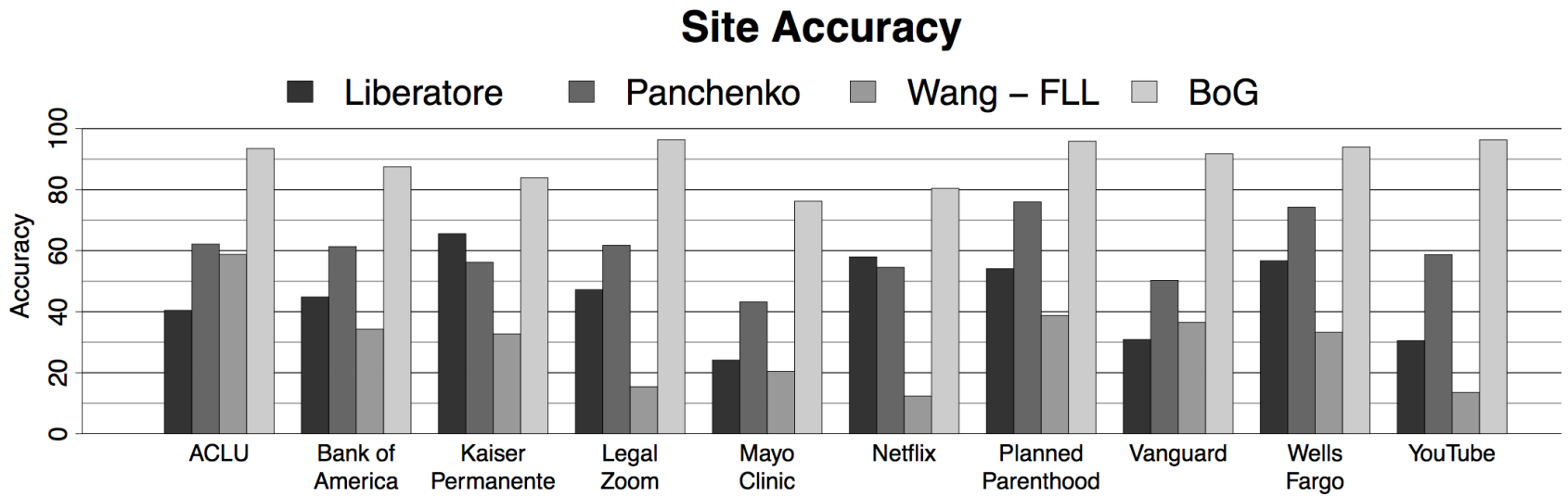
- Interactive pages that are responsive to user actions such as choices in drop-down menus, mouse clicks, typing

- Examples: healthcare diagnosis, taxation, web search (auto-complete)

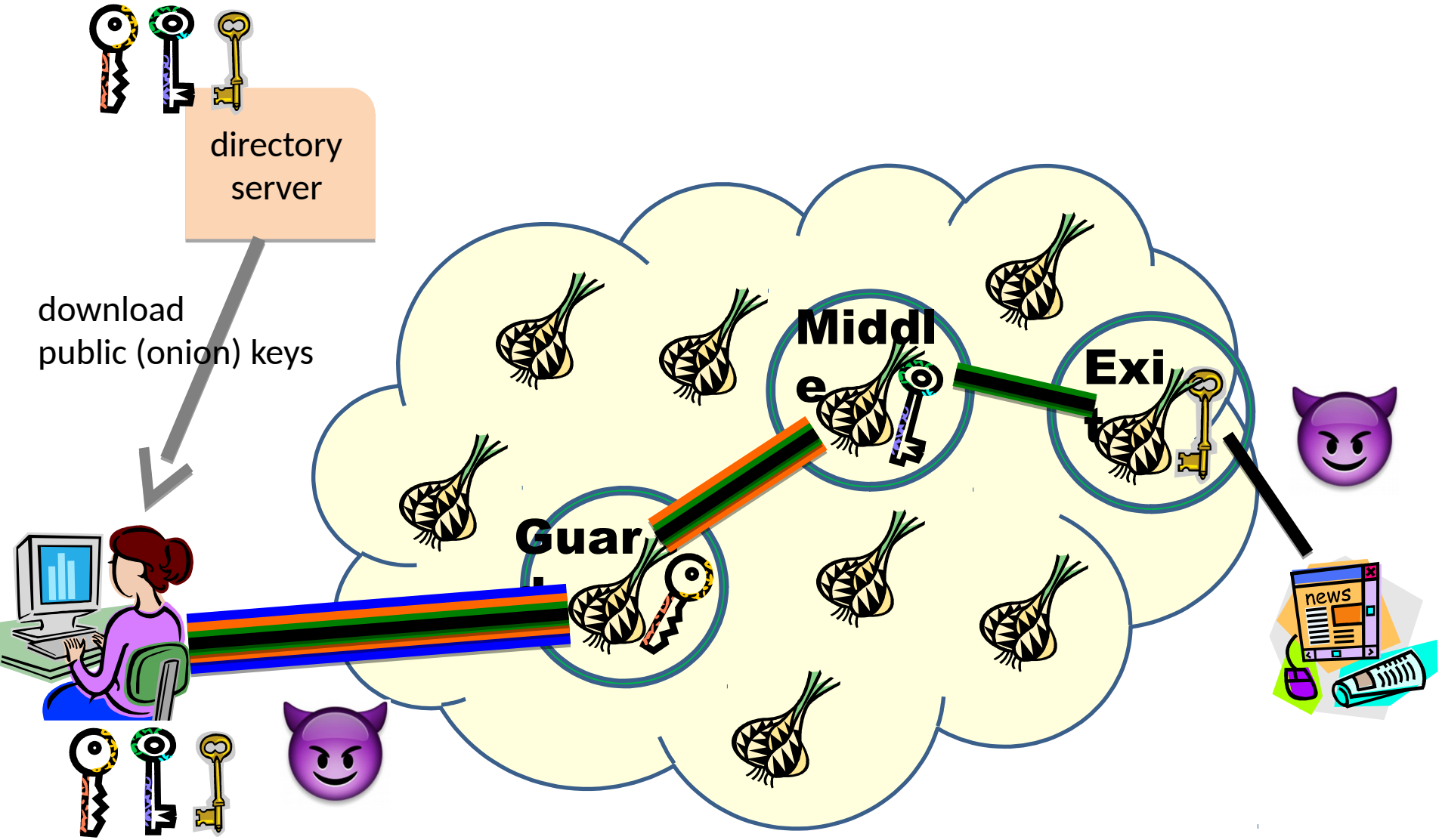
- Characteristics:
  - Stateful communication: transitions to next states depend both on the current state and on its input
  - Low entropy input: small input space
  - Uniqueness of traffic: disparate sizes and patterns for each possibility

“I know why you went to the clinic” (Miller et al, 2014)

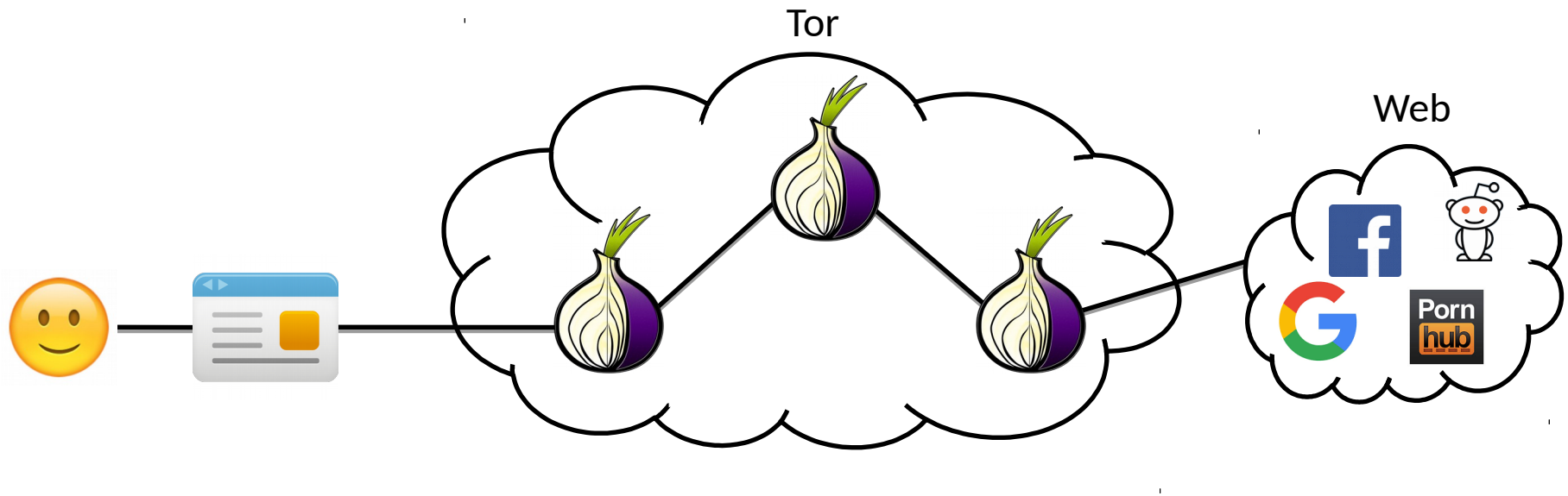
- Hidden Markov Models used to leverage link structure in websites
- Impact of caching and cookies was 17% (train with one option, test with the other)



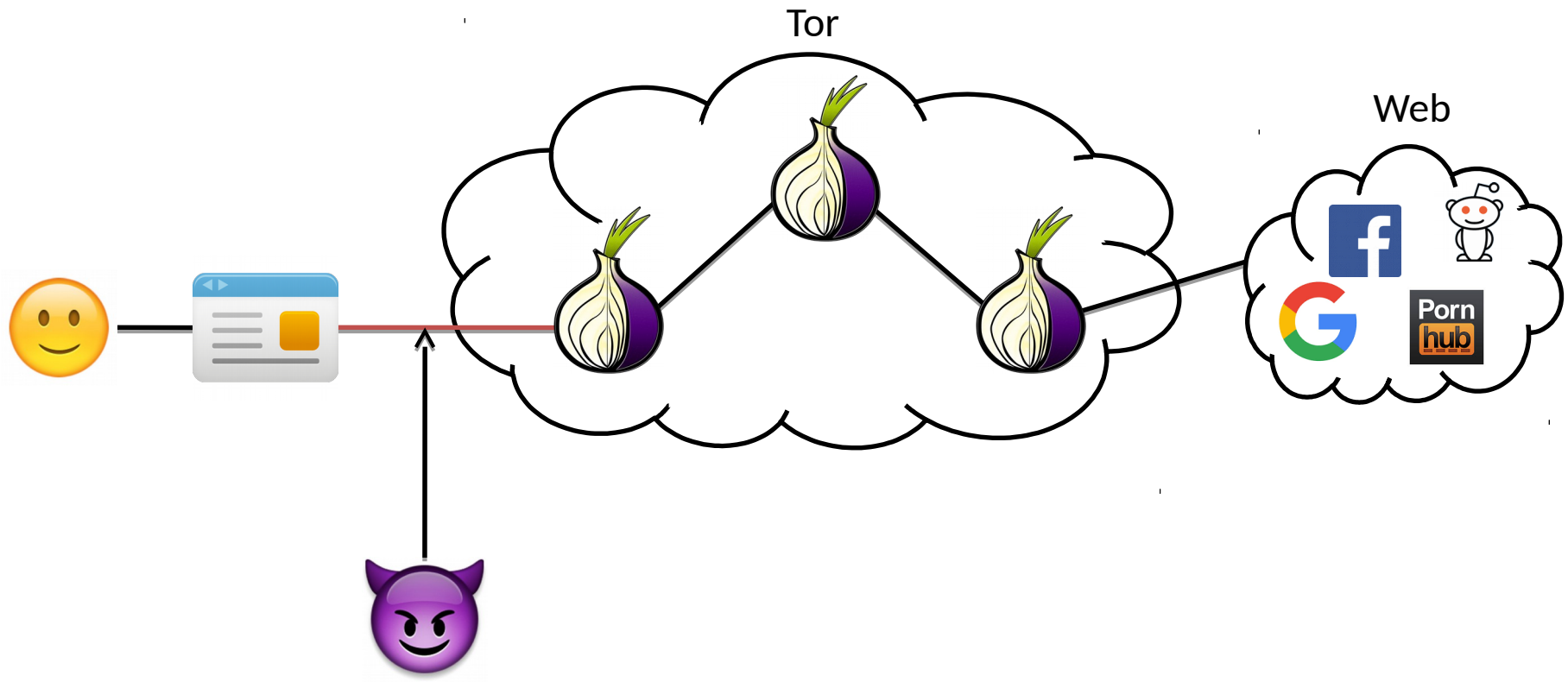
Tor


directory server
download public (onion) keys
Middl e
Exi
Guar
news

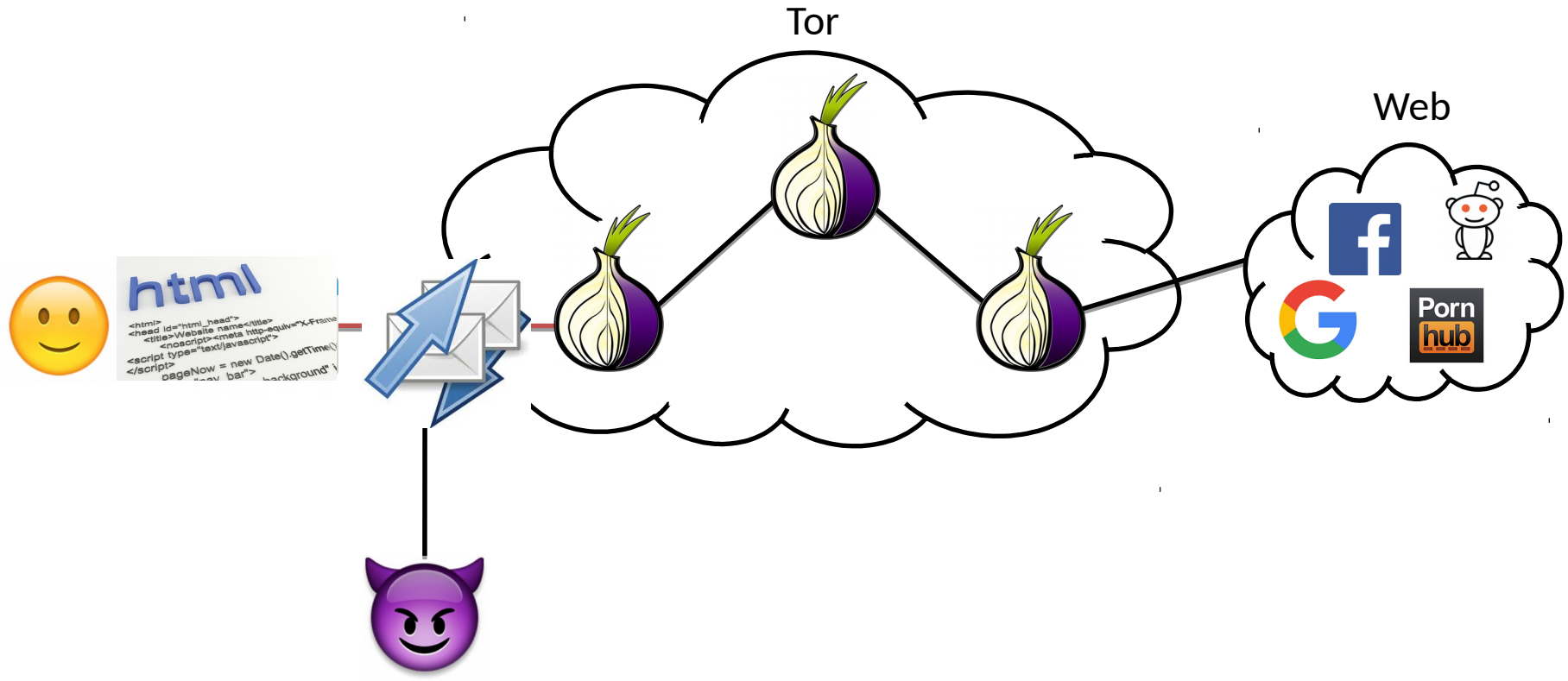
Tor



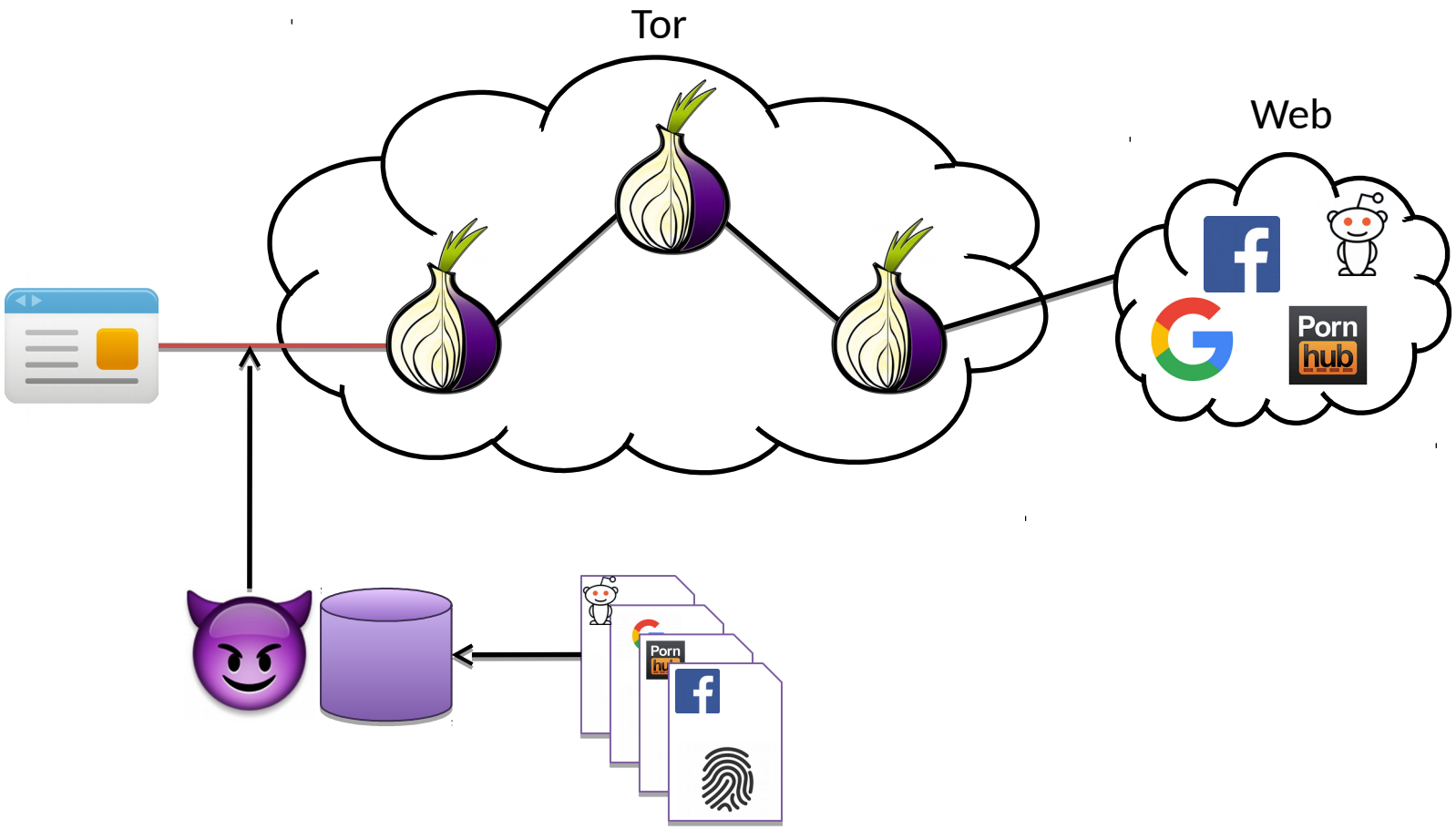
Website Fingerprinting



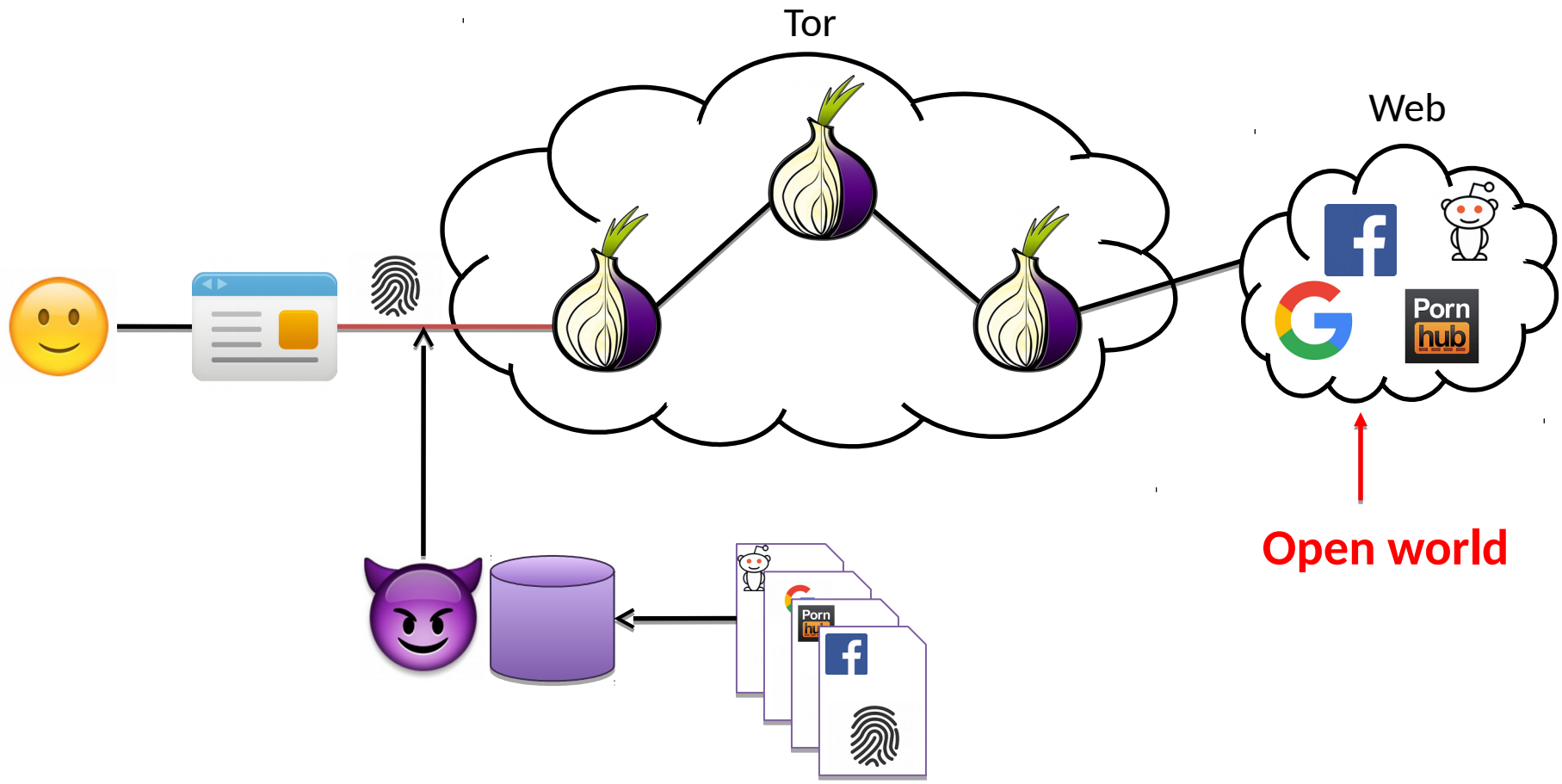
Website Fingerprinting



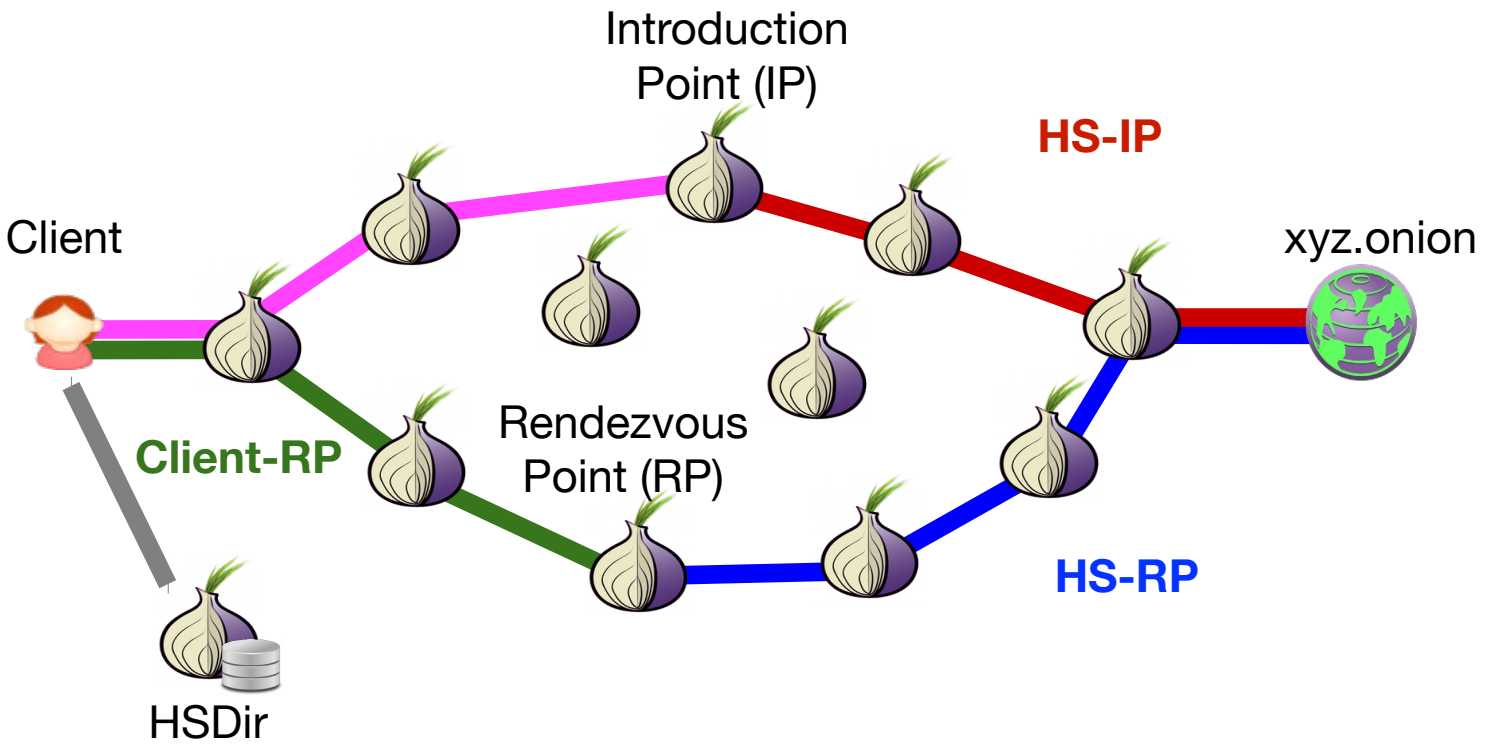
Website Fingerprinting



Website Fingerprinting



Tor Hidden (“Onion”) Services (HS)



HS-RP circuits are distinguishable from normal circuits (Kwon et al, 2015)

Size of the HS world is estimated at a few thousands (closed world!)

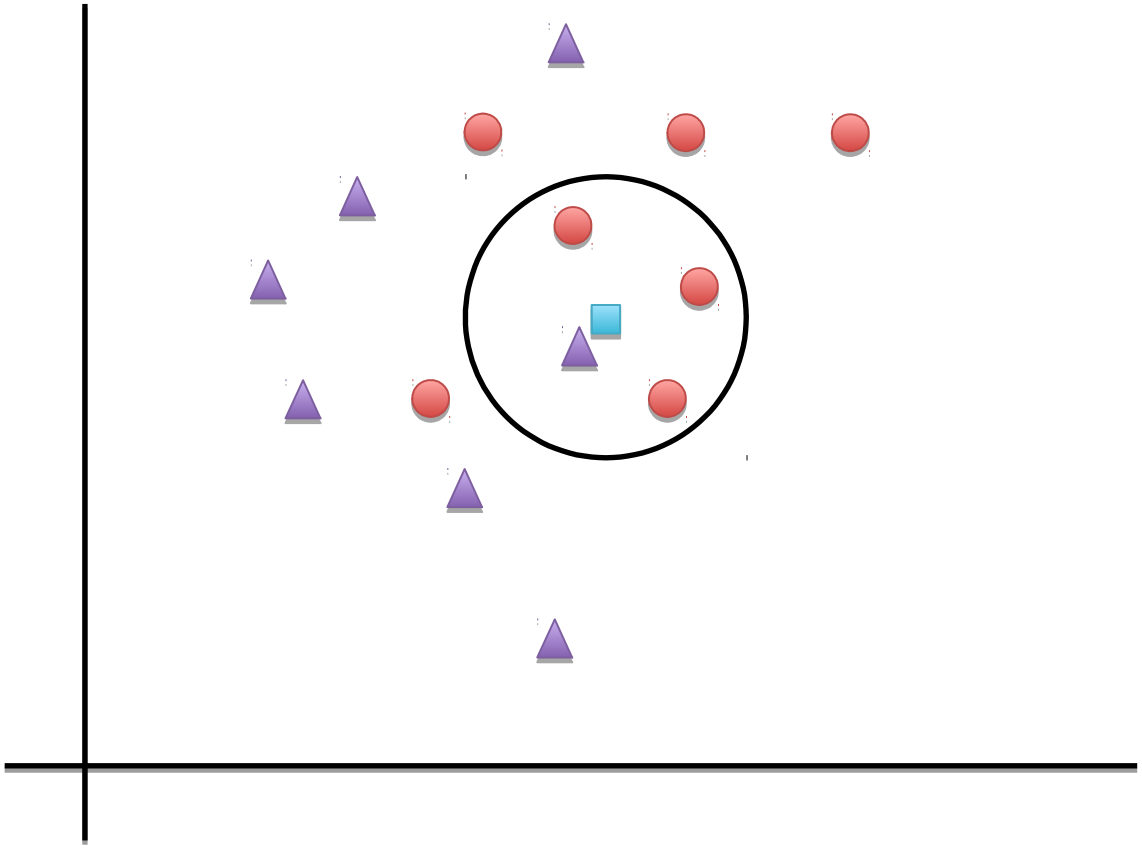
State of the art attacks

- kNN
- CUMUL
- k-Fingerprinting

kNN classifier (Wang et al, 2014)

- Features
  - 3,000
  - total size, total time, number of packets, packet ordering
  - the lengths of the first 20 packets
  - traffic bursts (sequences of packets in the same direction)
- Classification
  - *k*-NN
  - Tune weights of the distance metric that minimizes the distance among instances that belong to the same site.
- Results
  - 90% - 95% accuracy on a closed-world of 100 non-onion service websites.

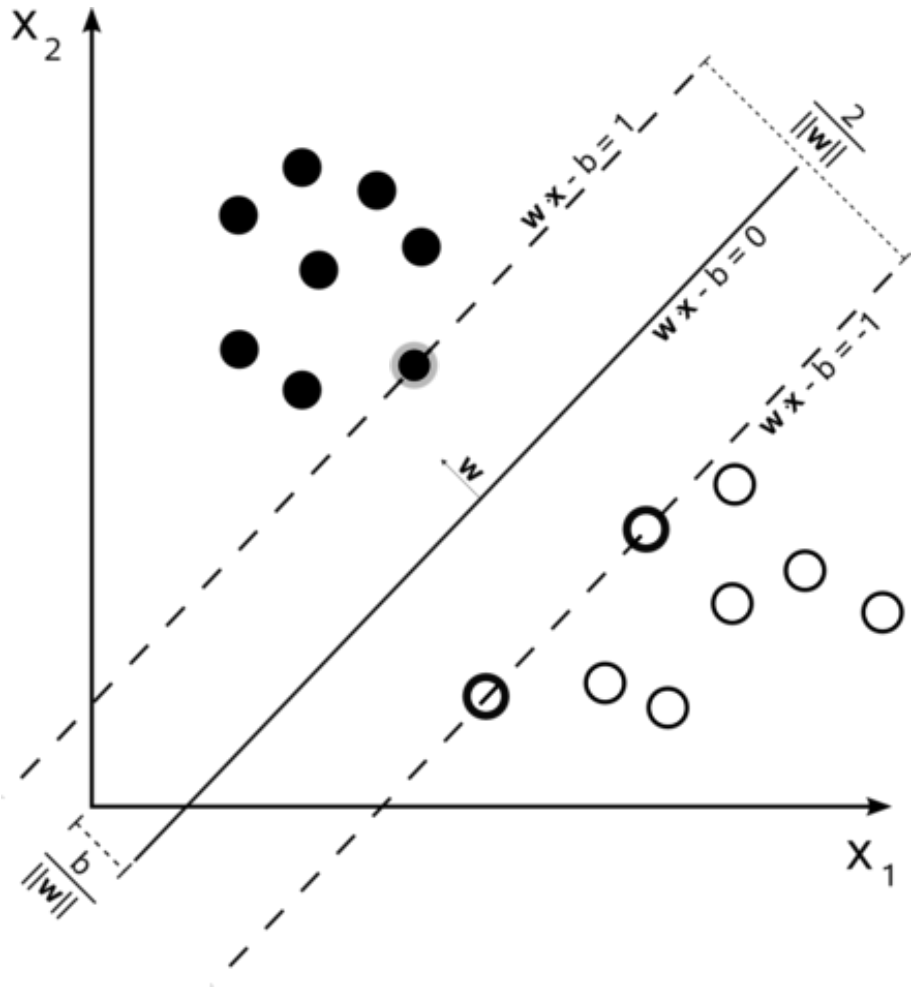
kNN



CUMUL (Panchenko et al, 2016)

- Features
  - a 104-coordinate vector formed by the number of bytes and packets in each direction and 100 interpolation points of the cumulative sum of packet lengths (with direction)
- Classification
  - Radial Basis Function kernel (RBF) SVM
- Results
  - 90% - 93% for 100 Non HS sites
  - Open world of 9,000 pages

SVM

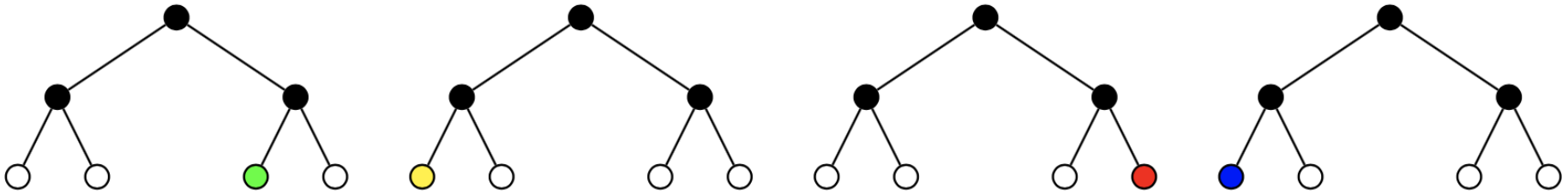


k-Fingerprinting (Hayes et al, 2016)

- Features
  - 175
  - Timing and Size features such as #packets/second
- Classification
  - Random Forest (RF) + k-NN
- Results
  - 90% accuracy on 30 onion services
  - Open world of 100,000 pages

Random Forest

- Train decision trees with web traffic features
- Training set is randomized per tree
- Random Forest is an ensemble of decision trees
- Use Random Forest output as the fingerprint of a website download

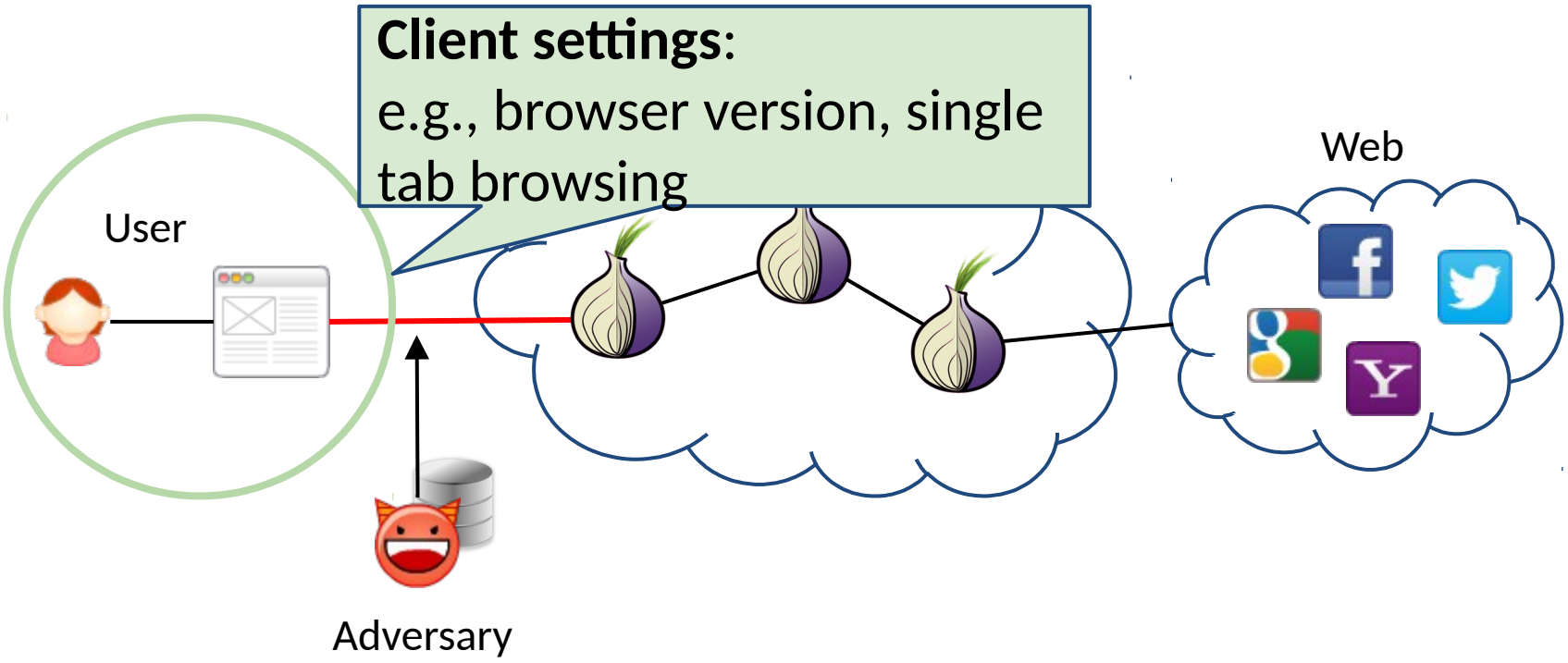


Why Do We Care?

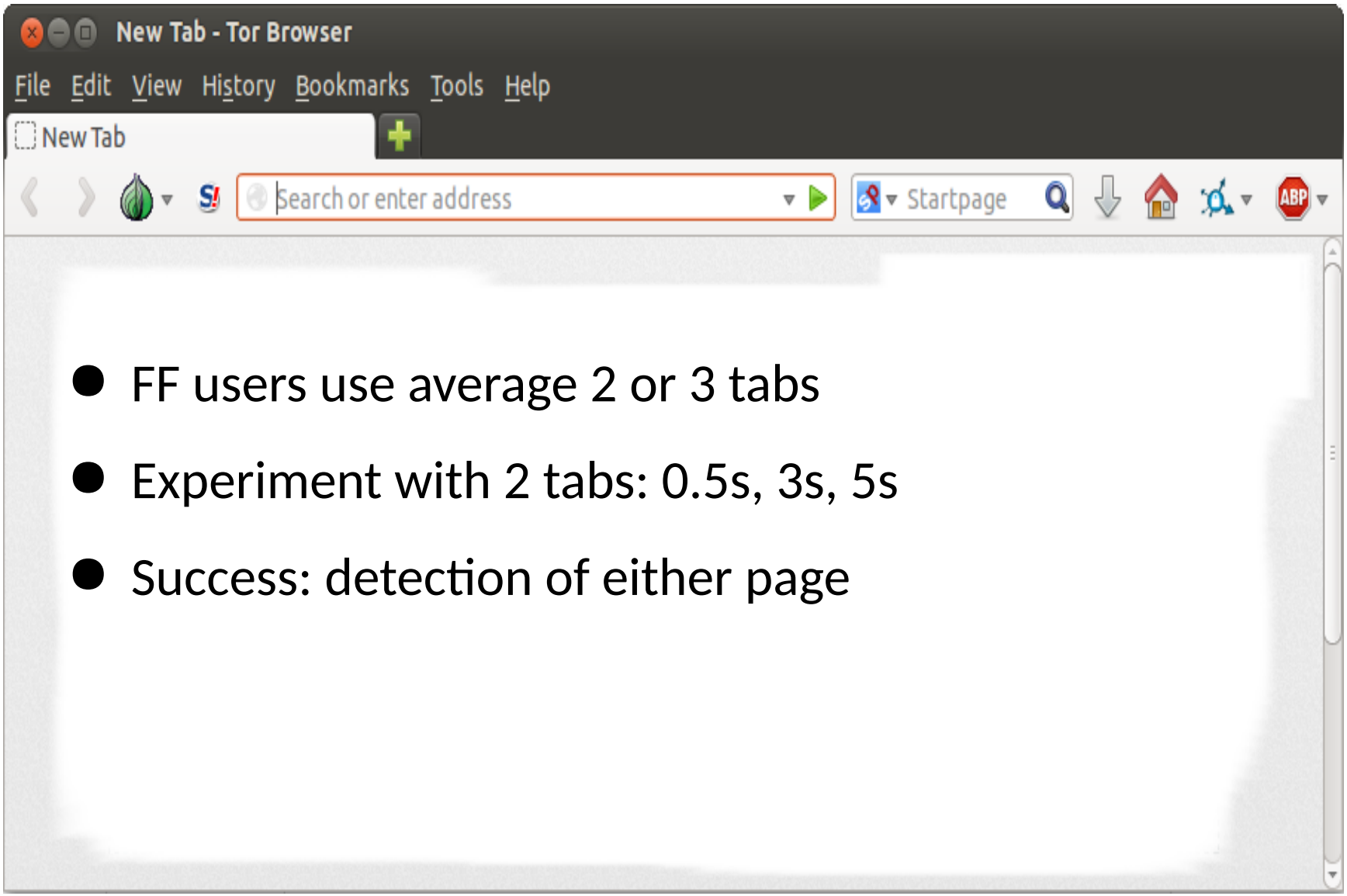
- Tor is the most advanced anonymity network
- WF allows an adversary to discover the browsing history
- Can be deployed by a low-resource adversary (that Tor aims to protect against)
- Series of successful attacks in the lab

- … how concerned should we be about these attacks *in practice*?
  - Critical review of WF attacks (Juarez et al, 2014)

Assumptions



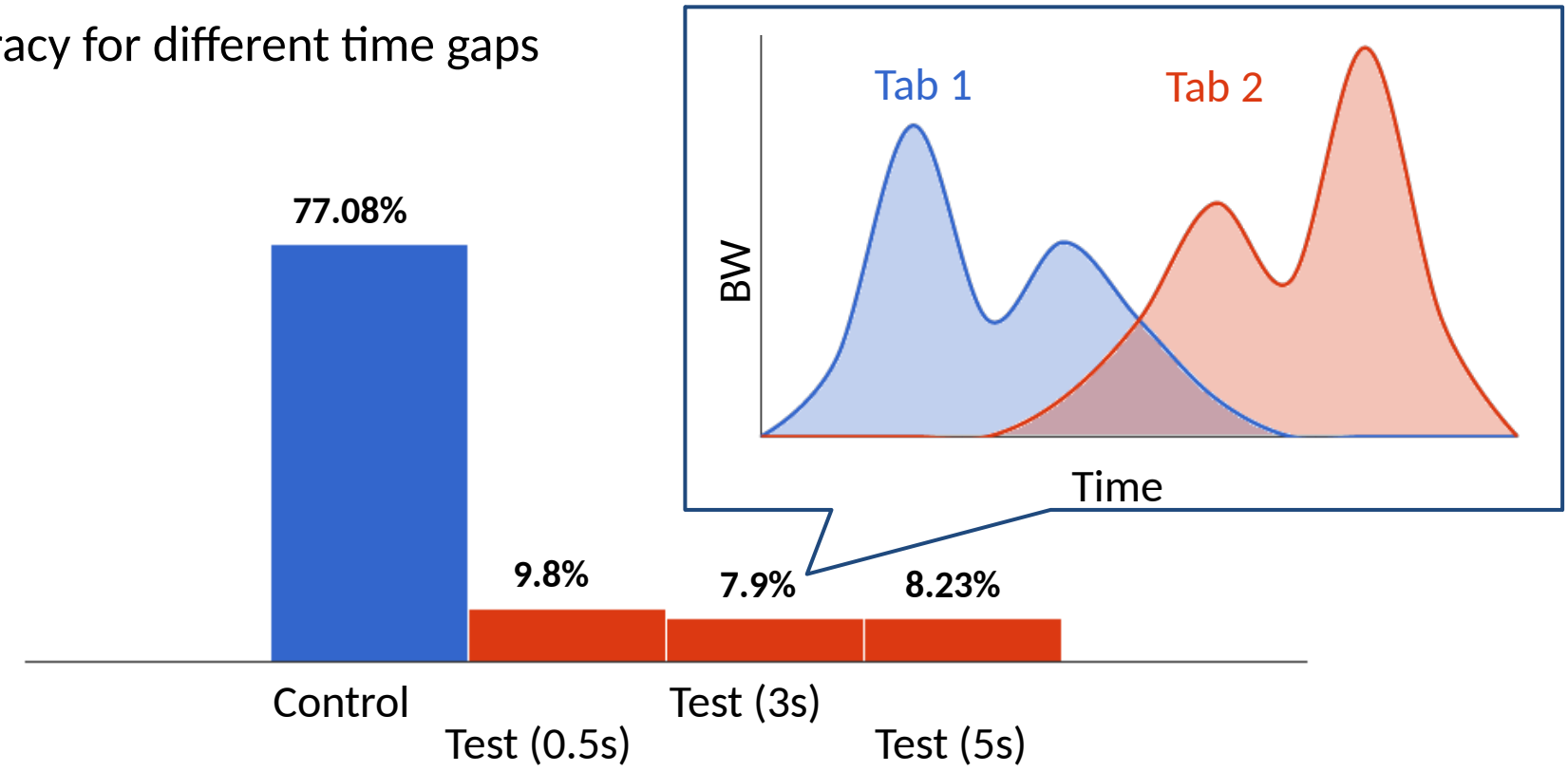
Effect of multi-tab browsing



- FF users use average 2 or 3 tabs
- Experiment with 2 tabs: 0.5s, 3s, 5s
- Success: detection of either page

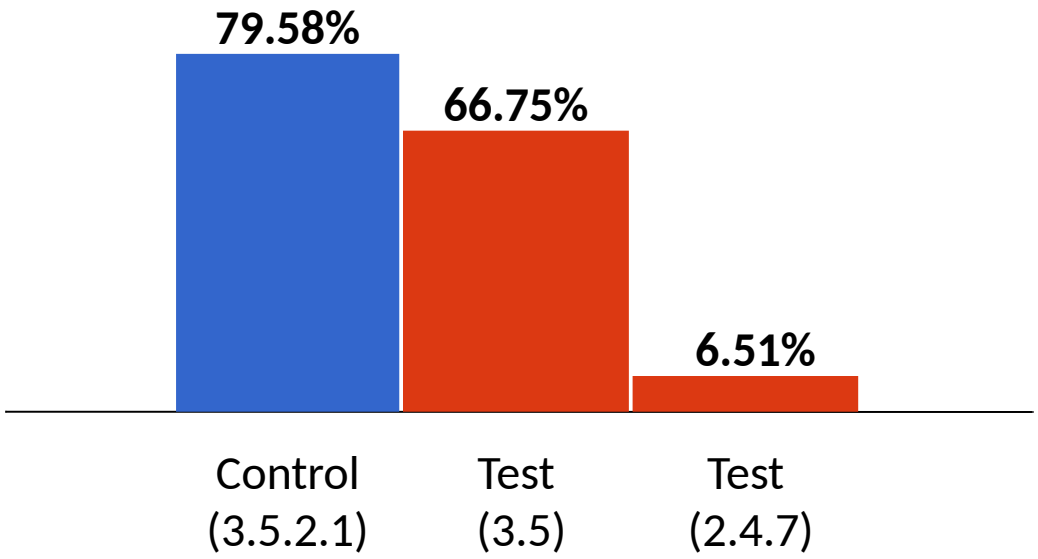
Experiments multi-tab

Accuracy for different time gaps

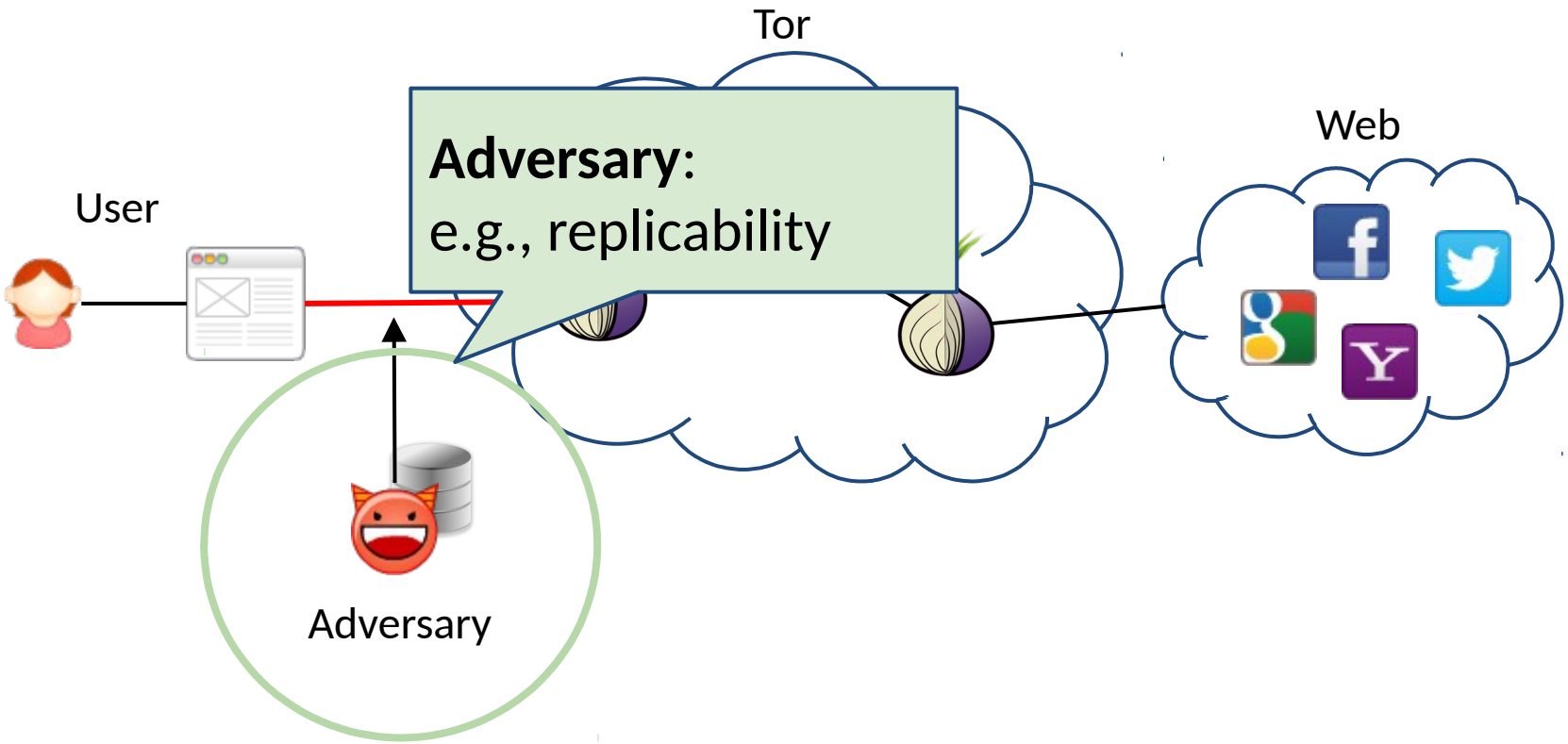


Experiments: TBB version

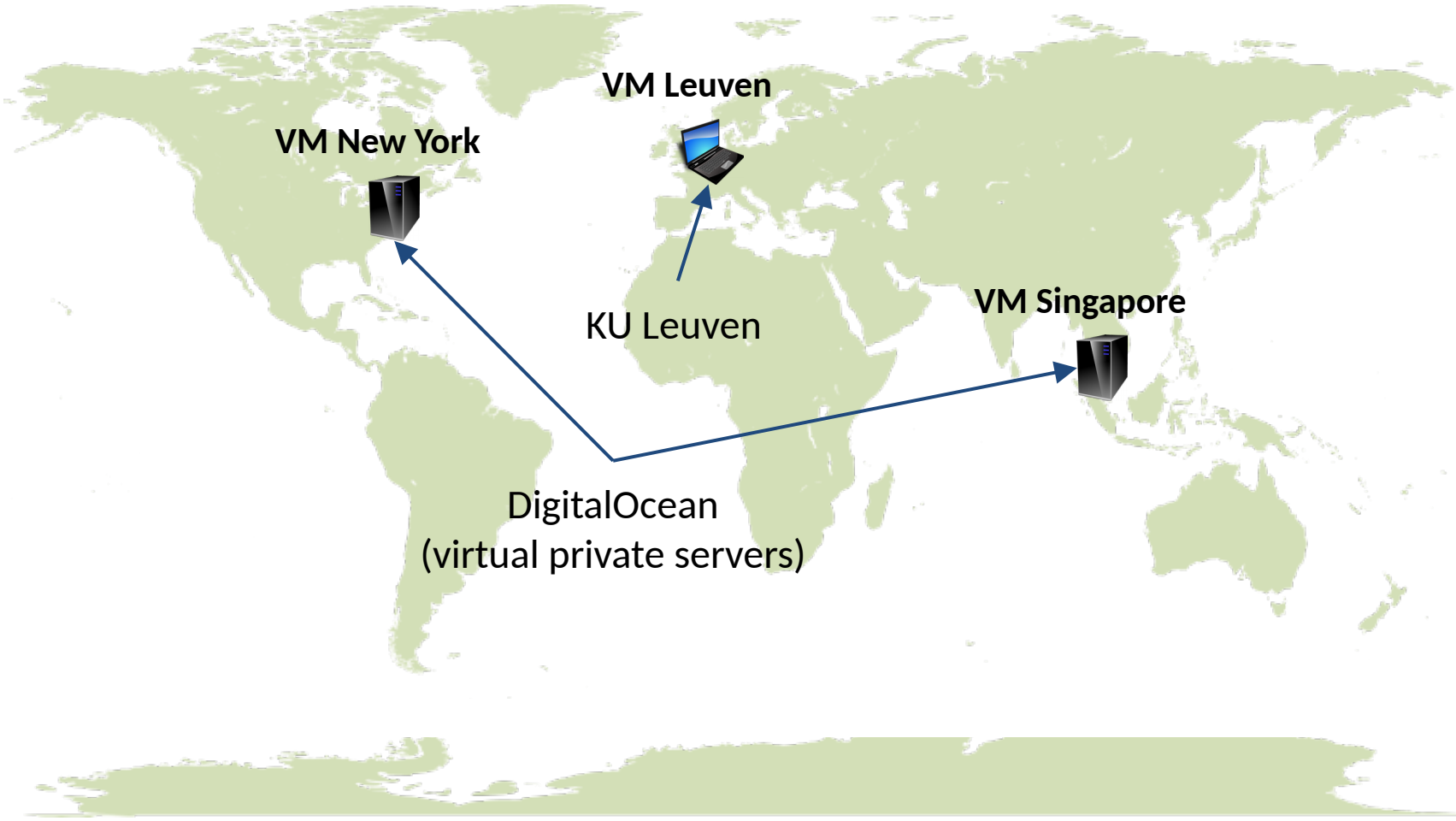
- TBB: Tor Browser Bundle
- Several versions coexist at any given time



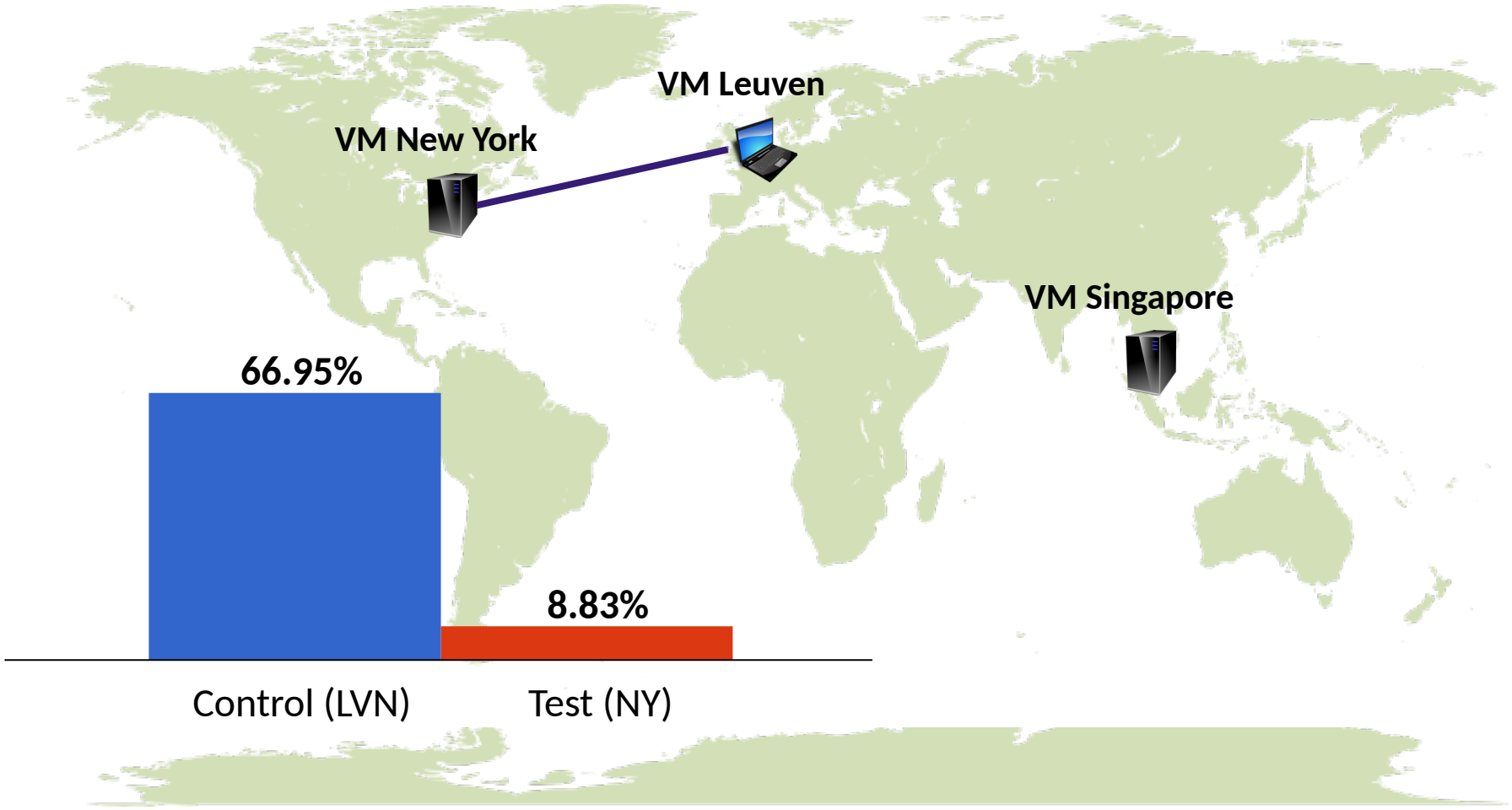
Assumptions



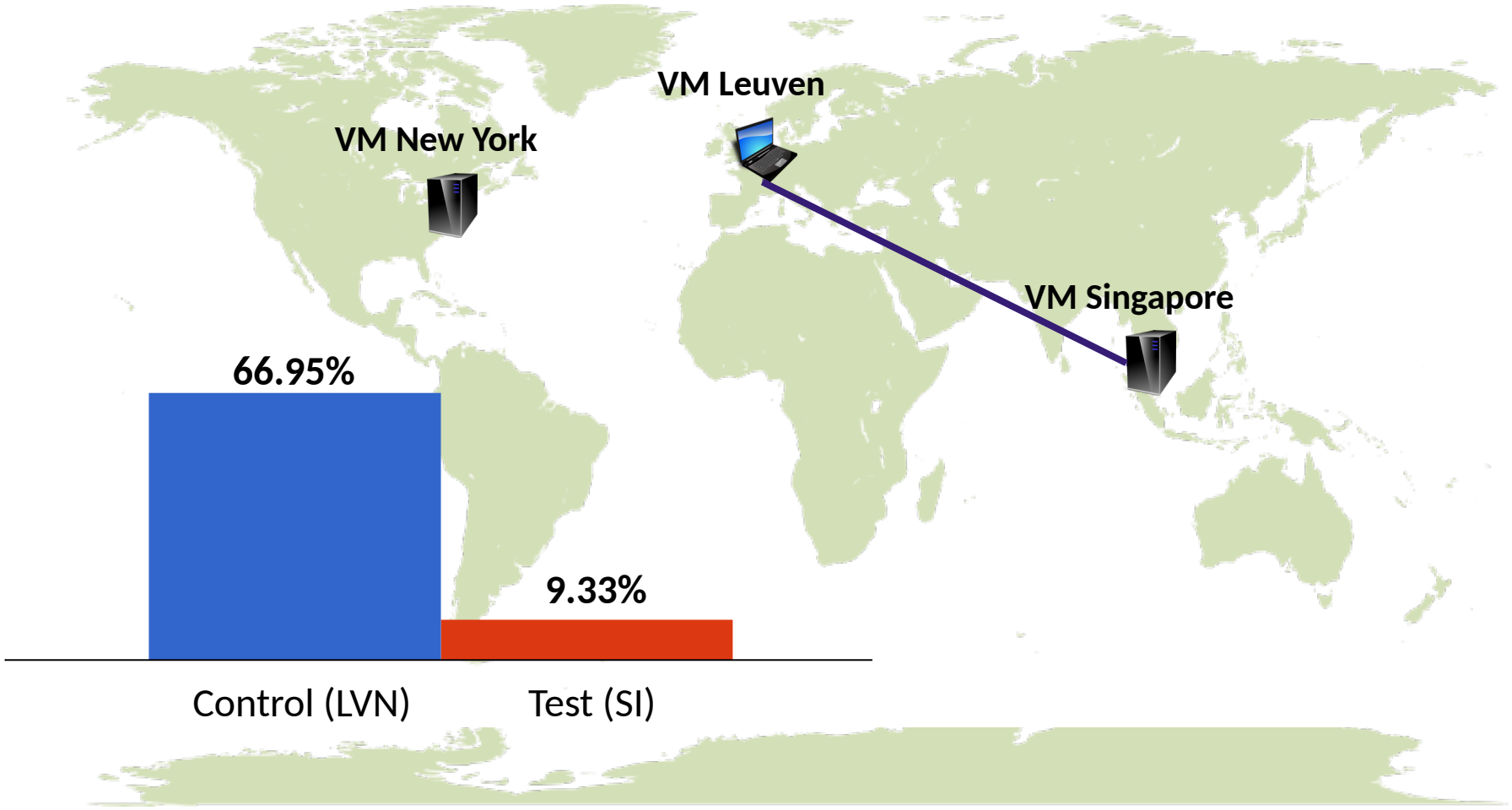
Experiments: network conditions



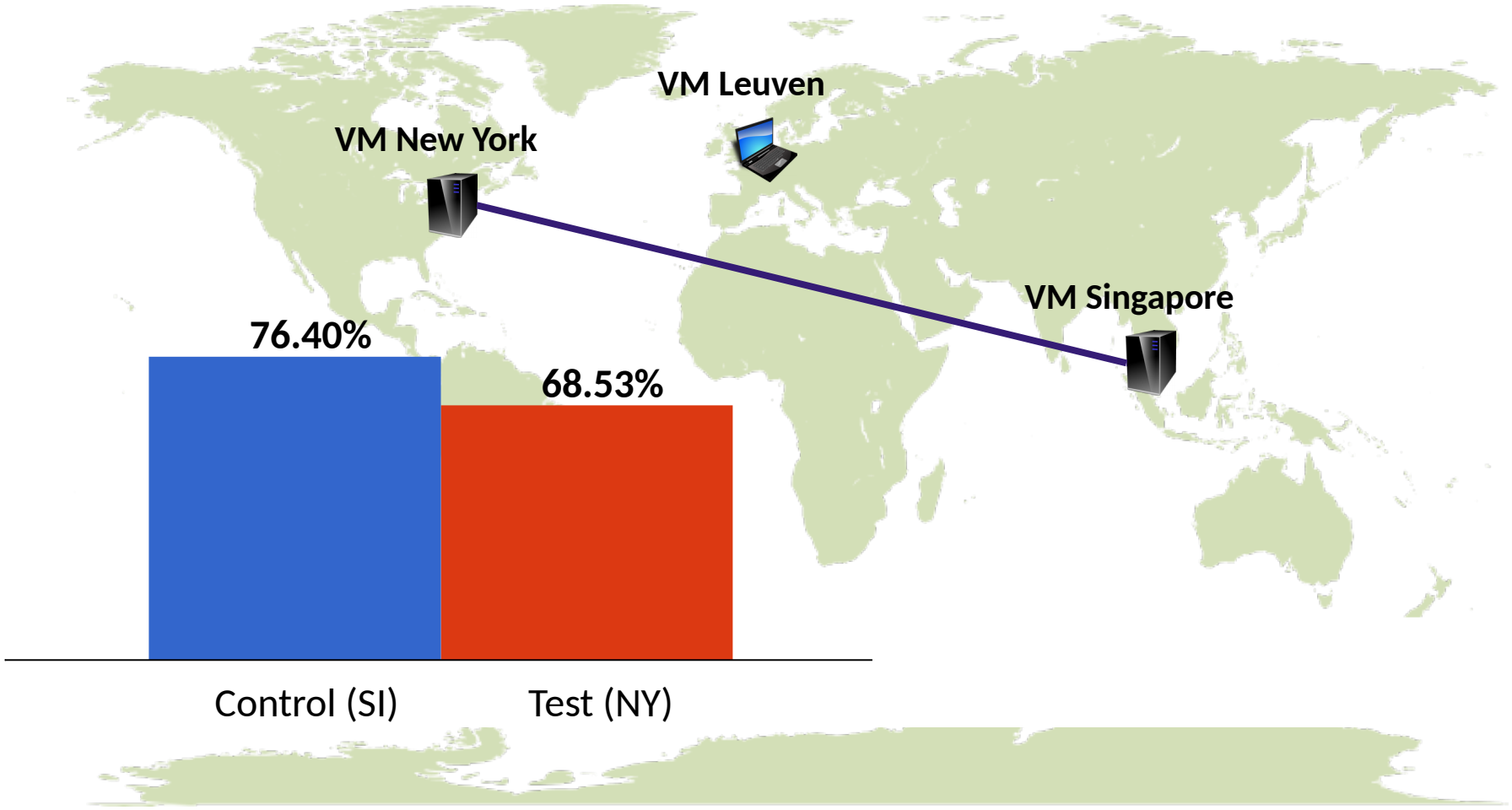
Experiments: network conditions



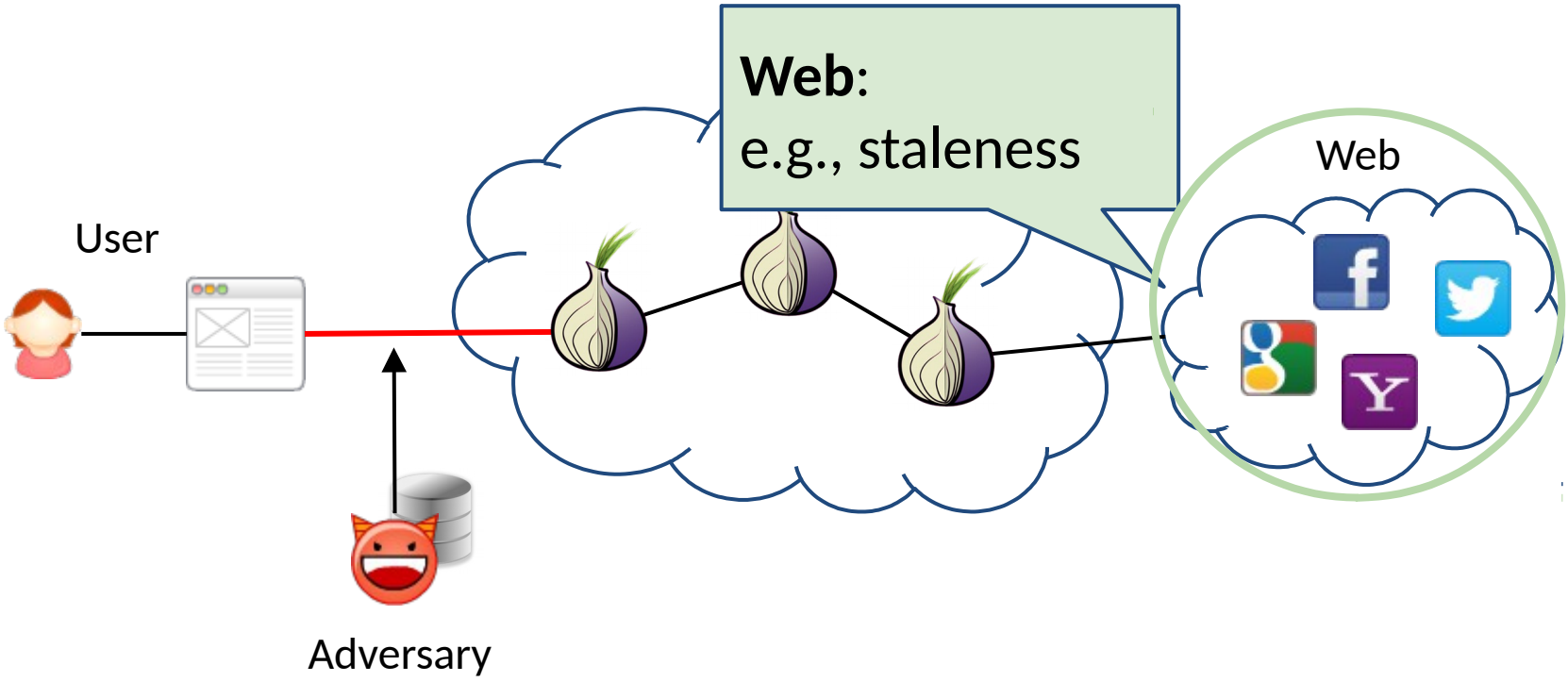
Experiments: network conditions



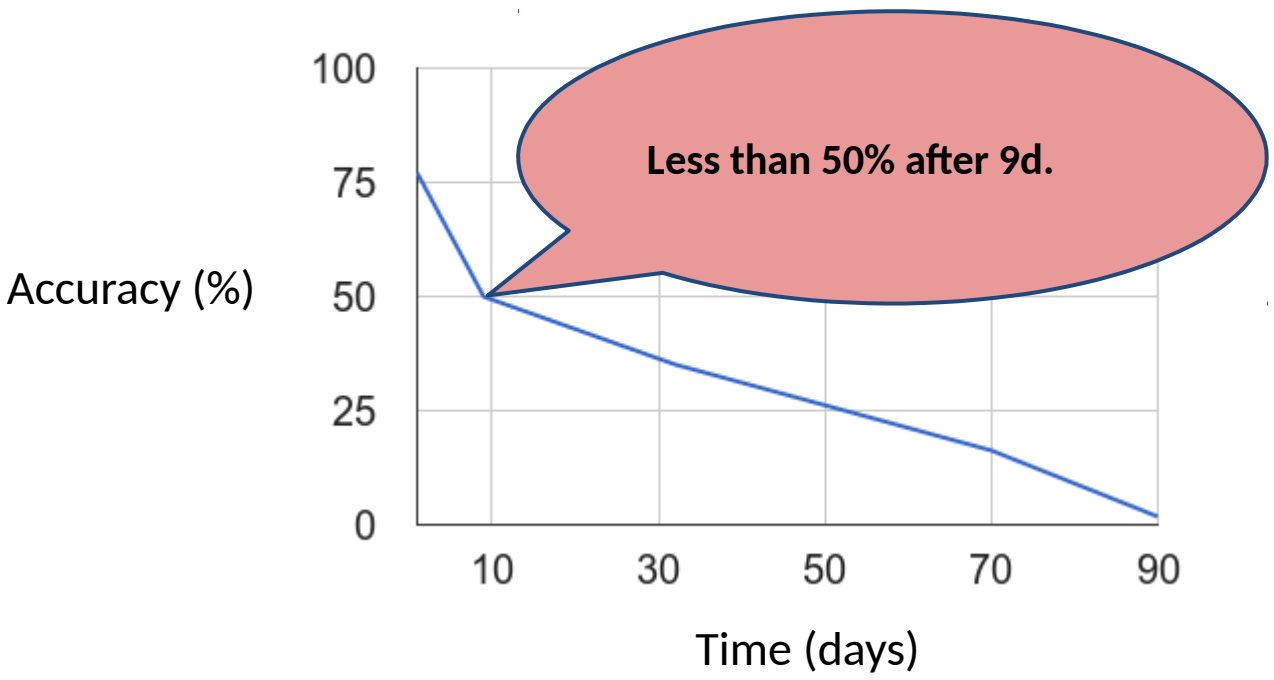
Experiments: network conditions



Assumptions



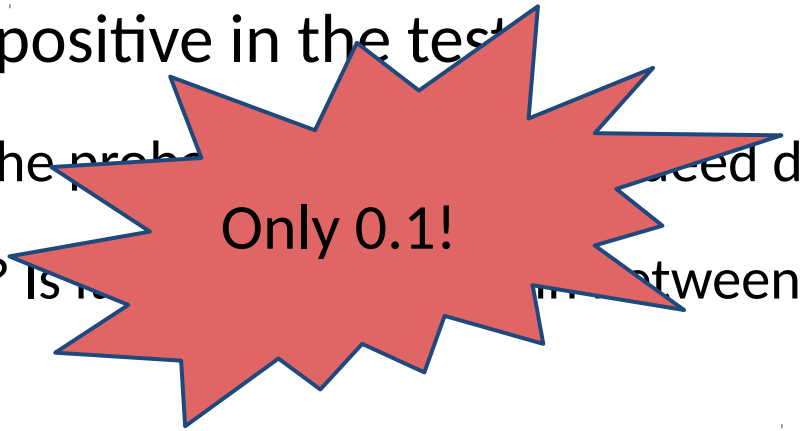
Data staleness



Effect of false negatives: Base rate fallacy

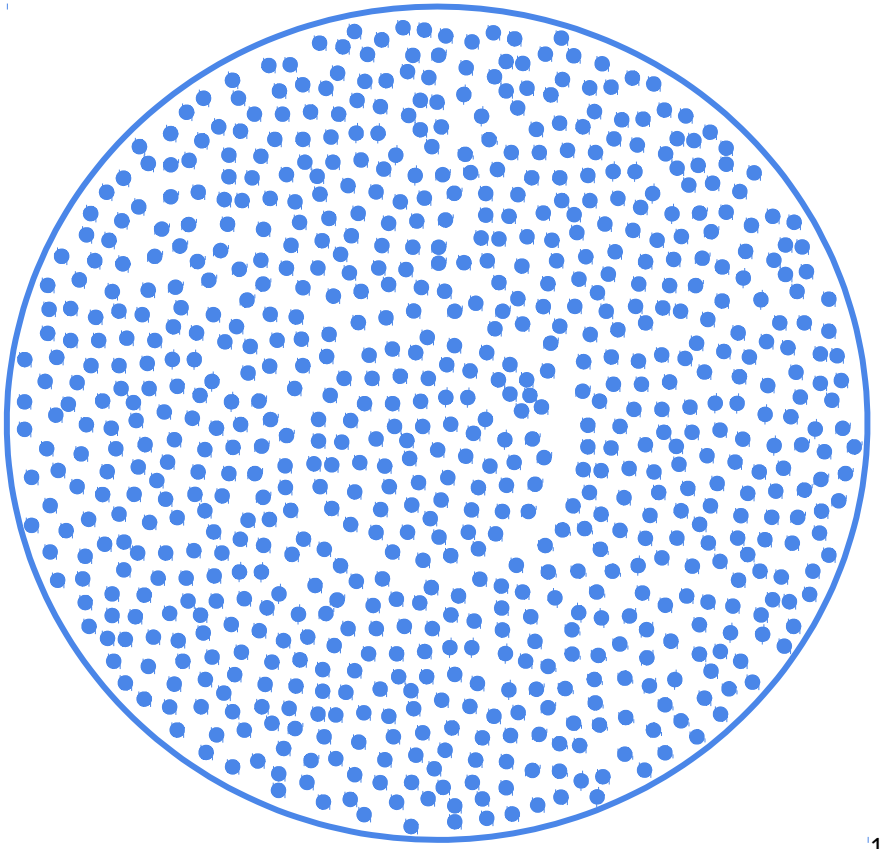
- Breathalyzer test:
  - **0.88** identifies truly drunk drivers (true positives)
  - **0.05** false positives

- Alice gives positive in the test
  - What is the prob… …ced drunk? (**BDR**)
  - Is it 0.95? Is it … …tween?



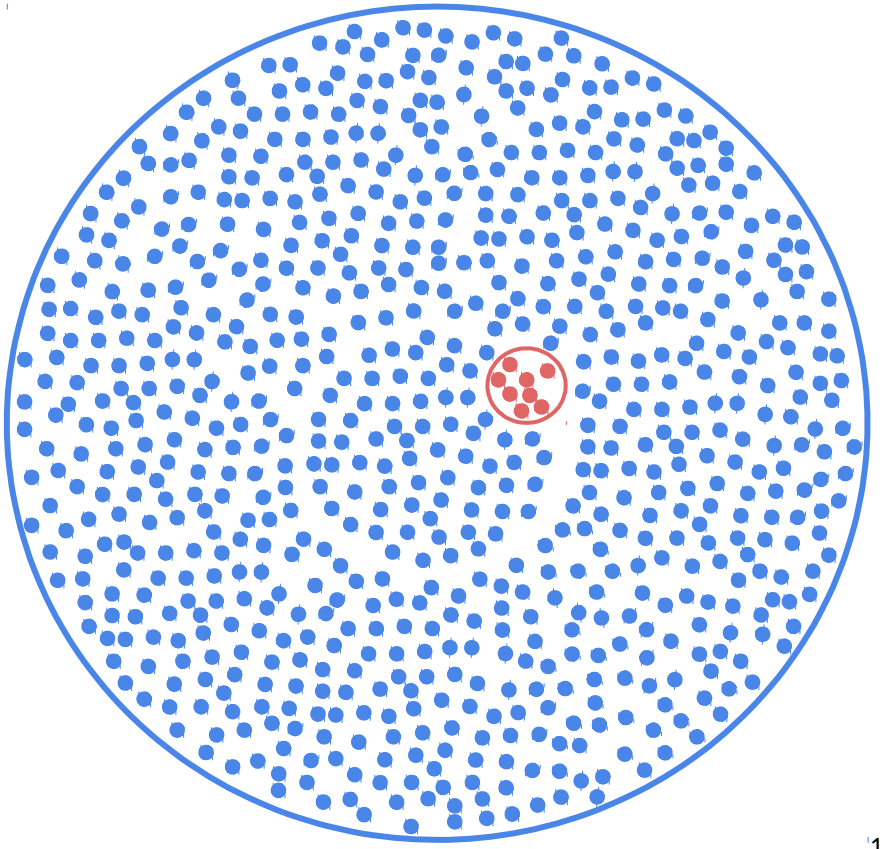
The base rate fallacy: example

- Circumference represents the world of drivers.
- Each dot represents a driver.



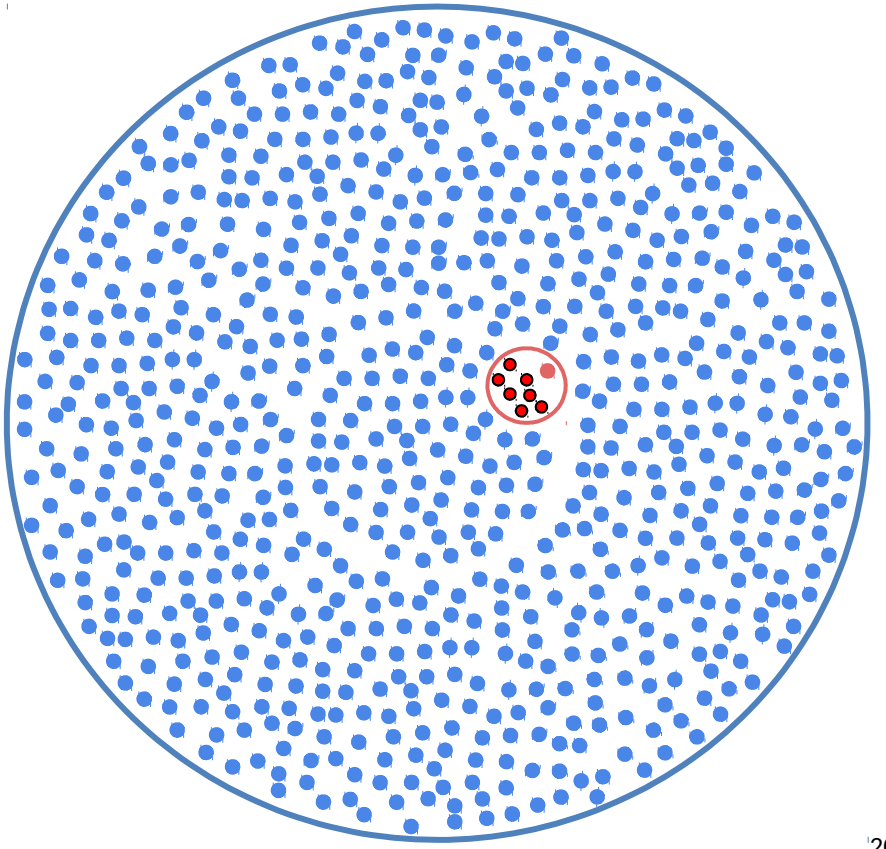
The base rate fallacy: example

- 1% of drivers are driving drunk (**base rate or prior**).



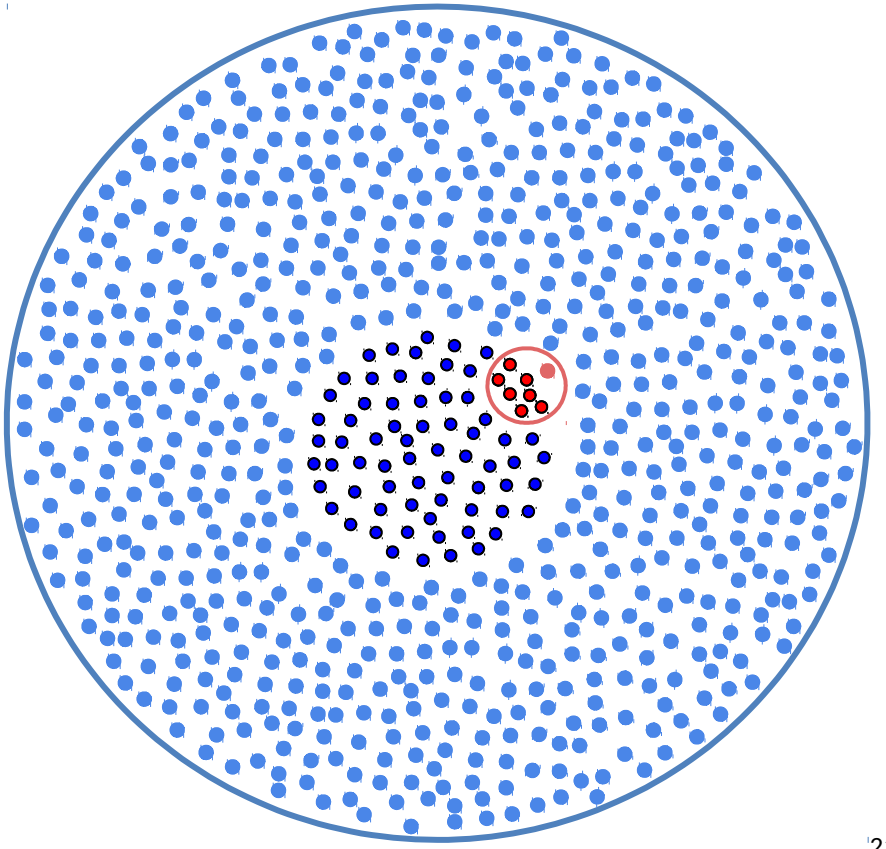
The base rate fallacy: example

- From drunk people 88% are identified as drunk by the test



The base rate fallacy: example

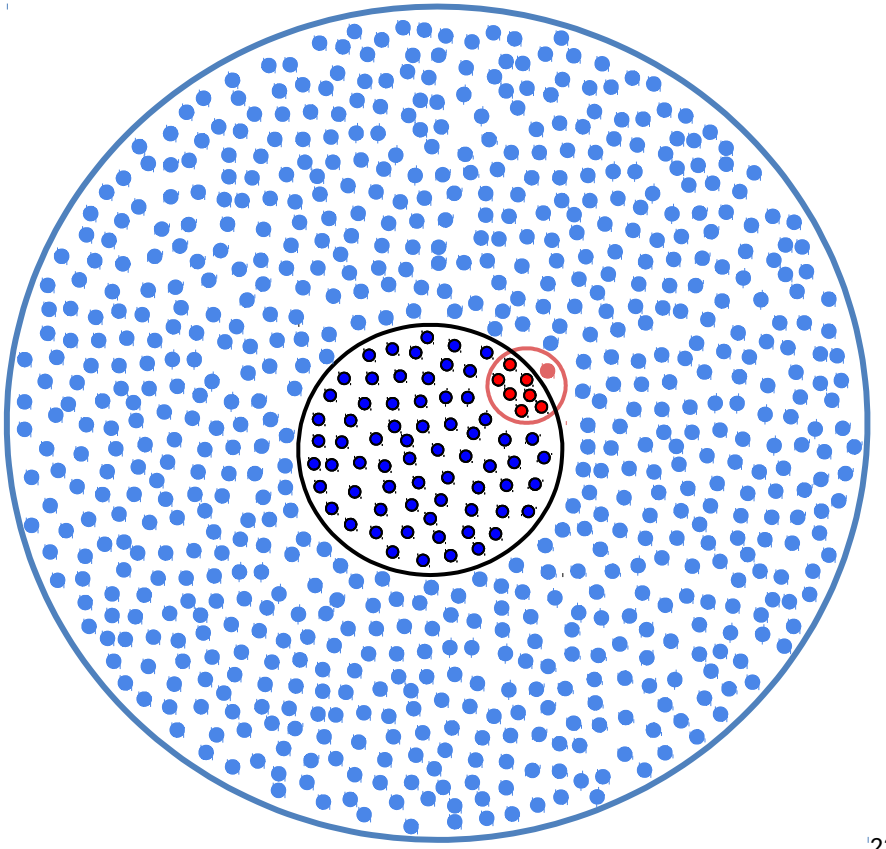
- From the not drunk people, 5% are erroneously identified as drunk



The base rate fallacy: example

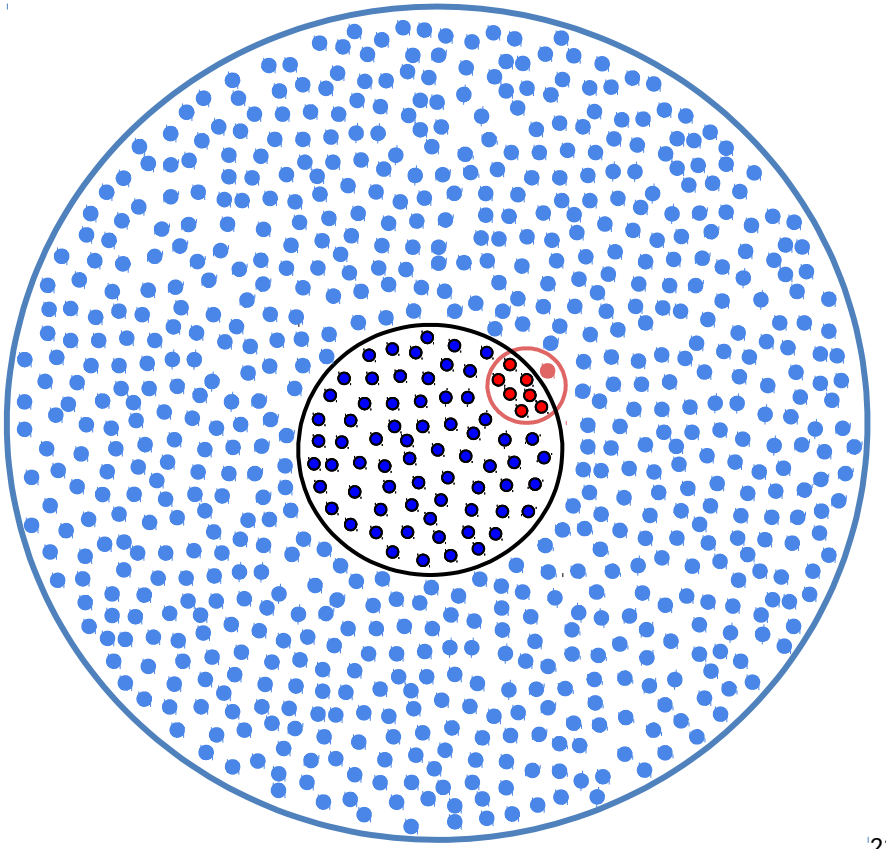
- Alice must be within the black circumference
- Ratio of red dots within the black circumference:

BDR = 7/70 = **0.1 !**



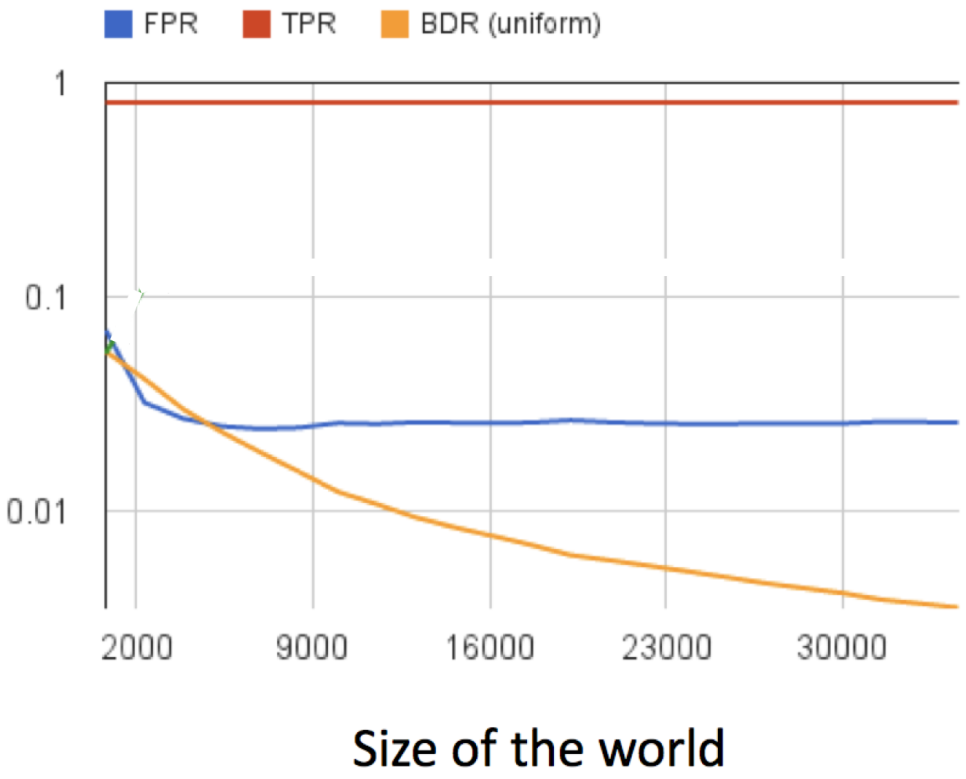
The base rate fallacy in WF

- Base rate must be taken into account
- In WF:
  - Blue: webpages
  - Red: monitored
  - Base rate?



Experiment: BDR in a 35K world

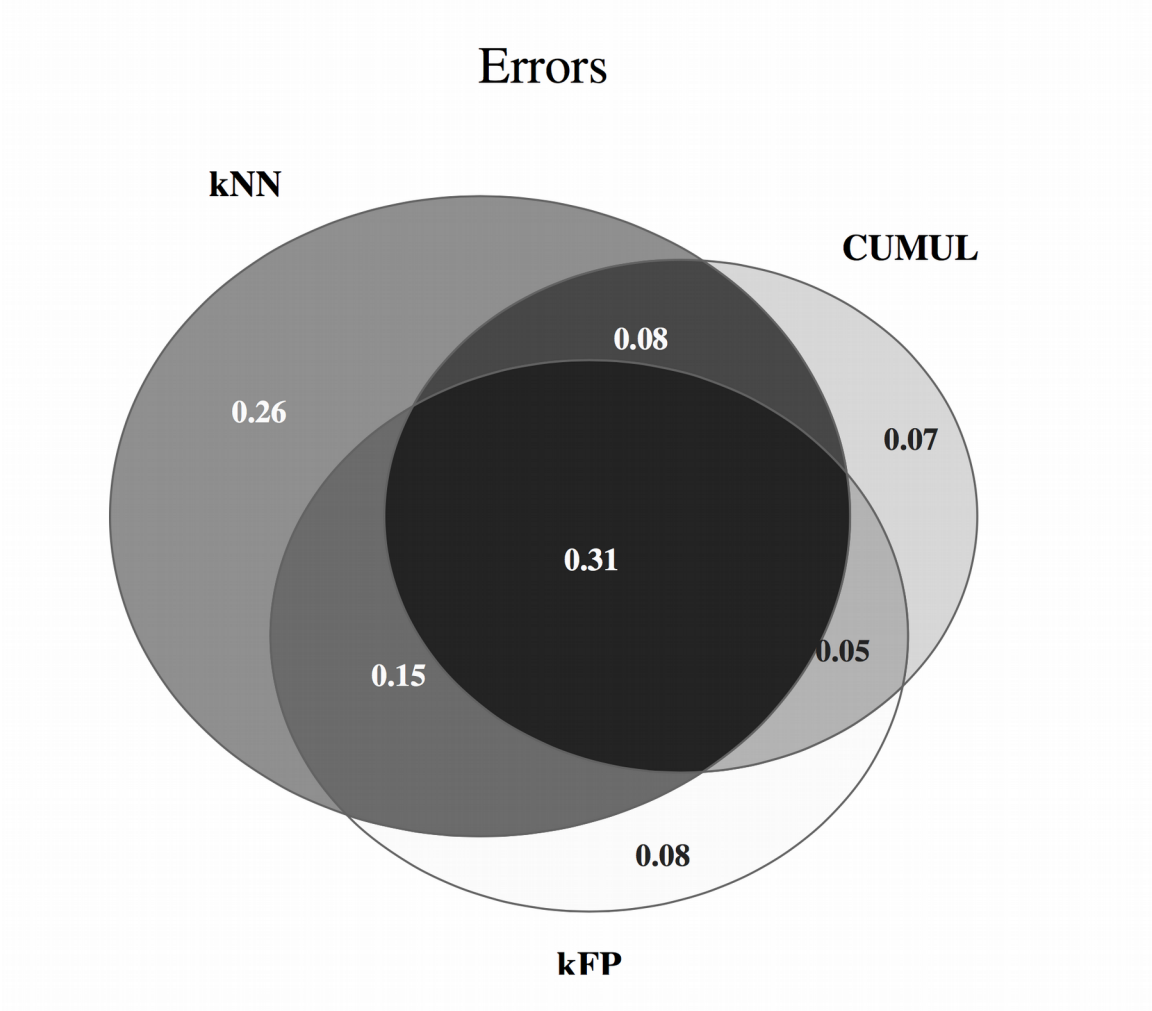
- World of 35K sites
- 4 target pages
- Uniform prior
- For 30K sites BDR is 0.4%



Disparate impact

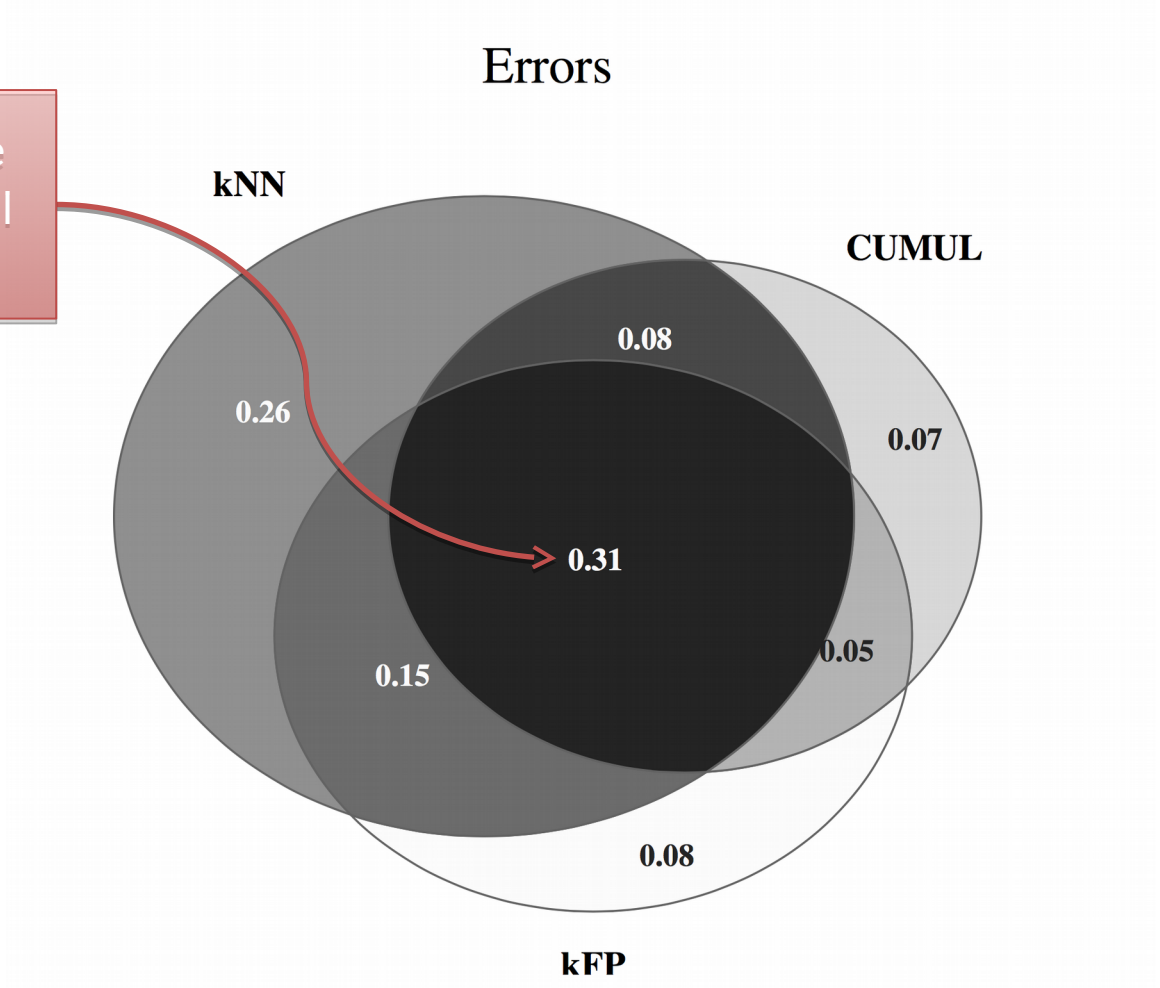
- WF normally attacks report average success
- But…
  - Are certain websites more susceptible to website fingerprinting attacks than others?
  - What makes some sites more vulnerable to the attack than others?

Misclassifications of onion services: Sites that are “safe”

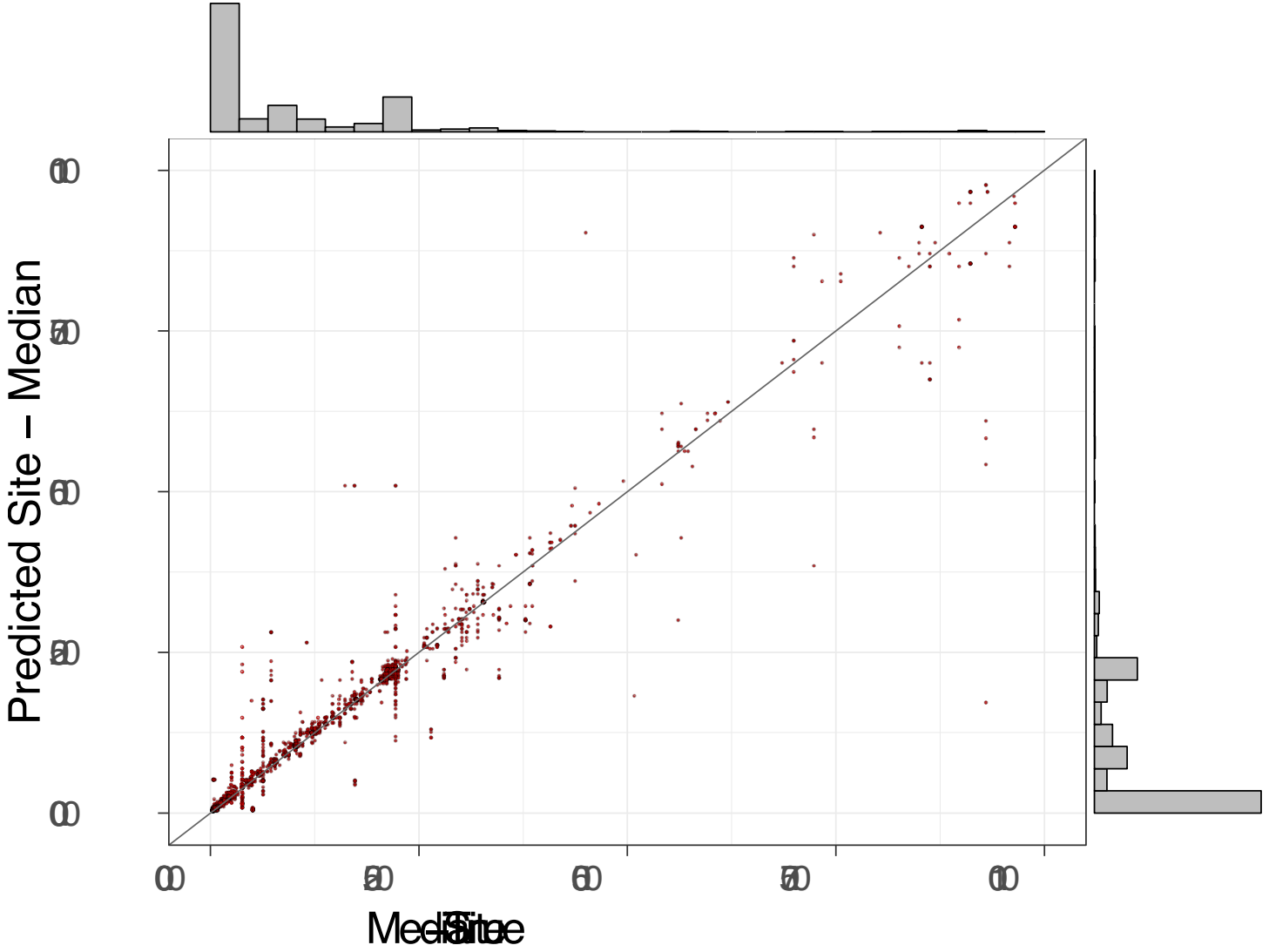


Misclassifications: Sites that are “safe”

Some sites are hidden from all methods!



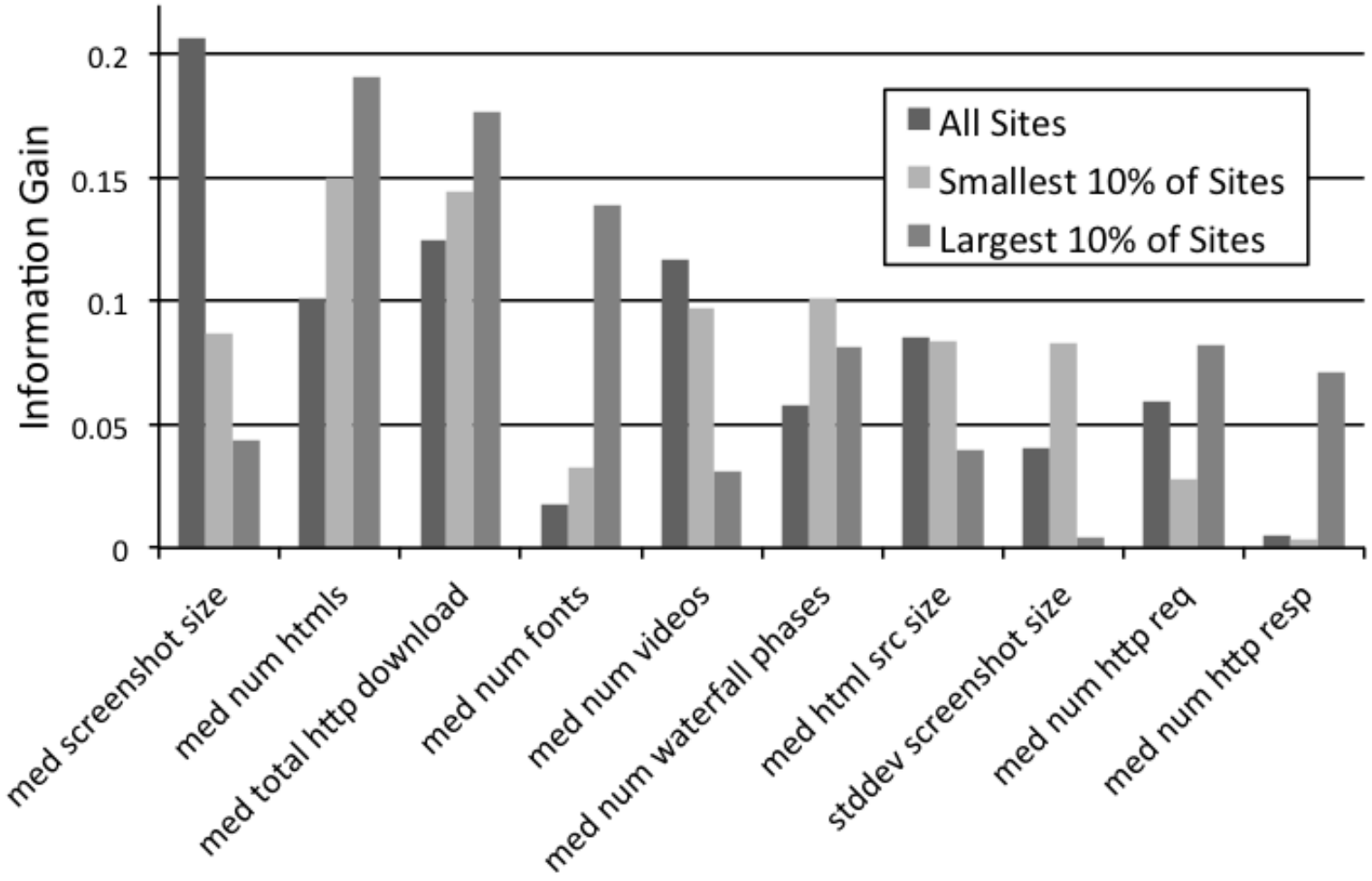
Median of total incoming packet size for misclassified instances



Site-level Feature Analysis

- Trace features are not always helpful
- Can we determine what characteristics of a website affect its fingerprintability?
- Site-Level Features:
  - Total HTTP download size
  - http duration
  - screenshot size
  - number of scripts
  - …

Site Level Feature Analysis



WF countermeasures

- Network layer
  - Add padding
    - Constant rate is unreasonable
    - Leakage: how to optimize padding?
  - Add latency to disrupt the traffic pattern
    - Bad idea
- Page design
  - Small size
  - Dynamism

To conclude

- WF can be deployed by adversaries with only local access to the communications network
- WF seriously undermines the protection offered by https
- WF threatens the anonymity properties of Tor
  - Though it's unclear to which extent lab results would hold in the wild
  - The attack is costly in terms of resources
- Disparate impact: some pages are more fingerprintable than others, which is not captured if you only look at average results
- Countermeasures involve additional traffic and/or dynamism