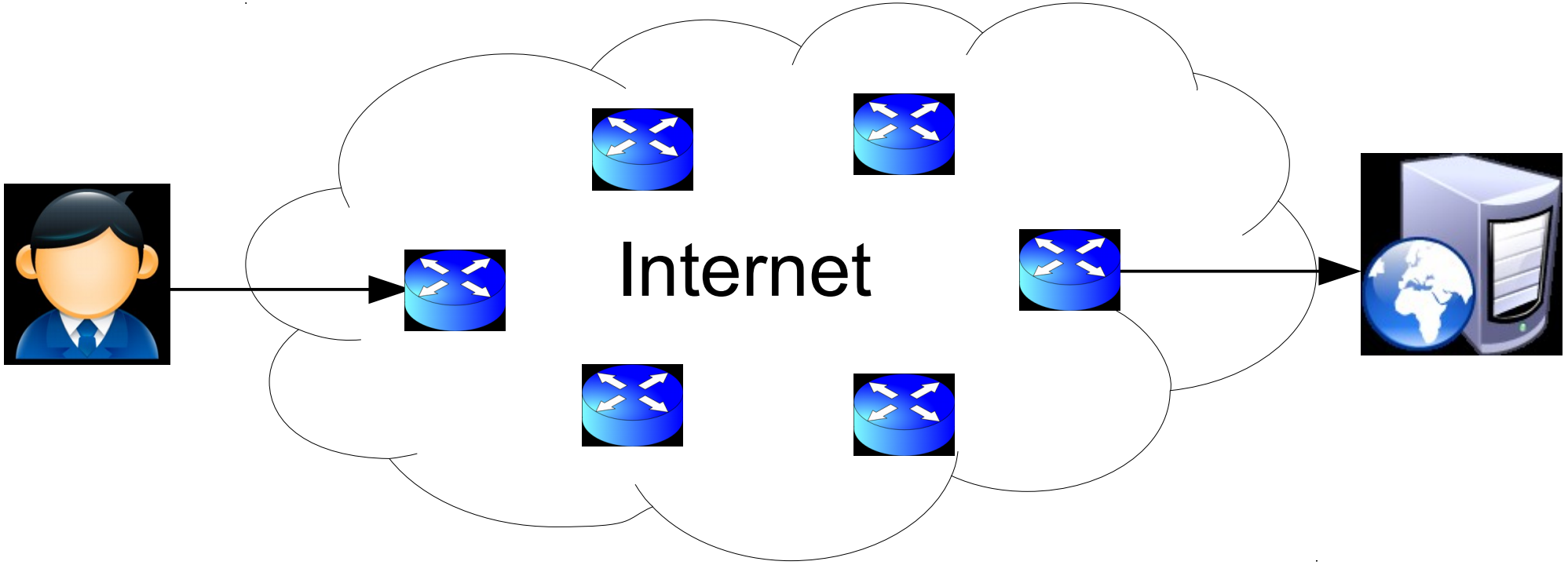
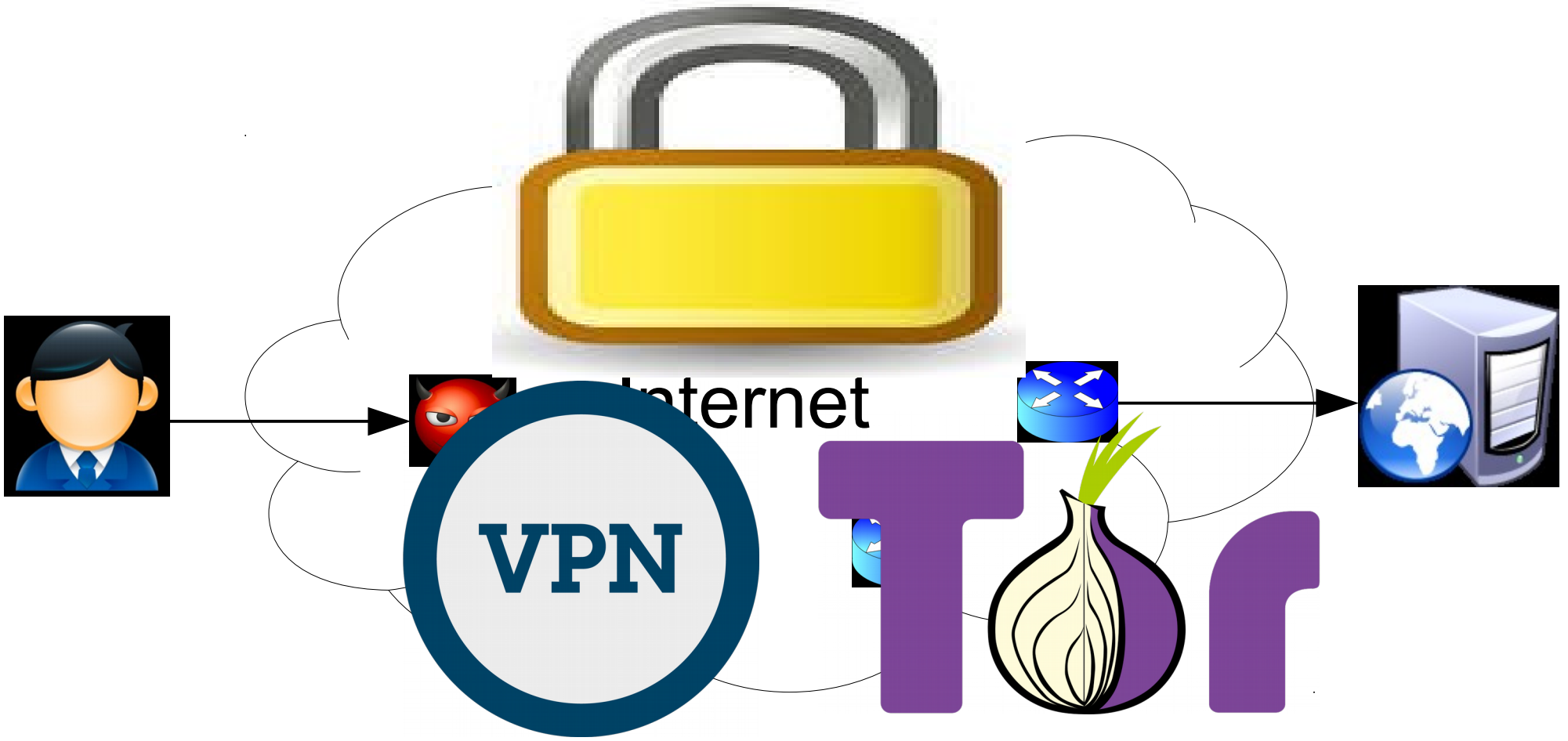
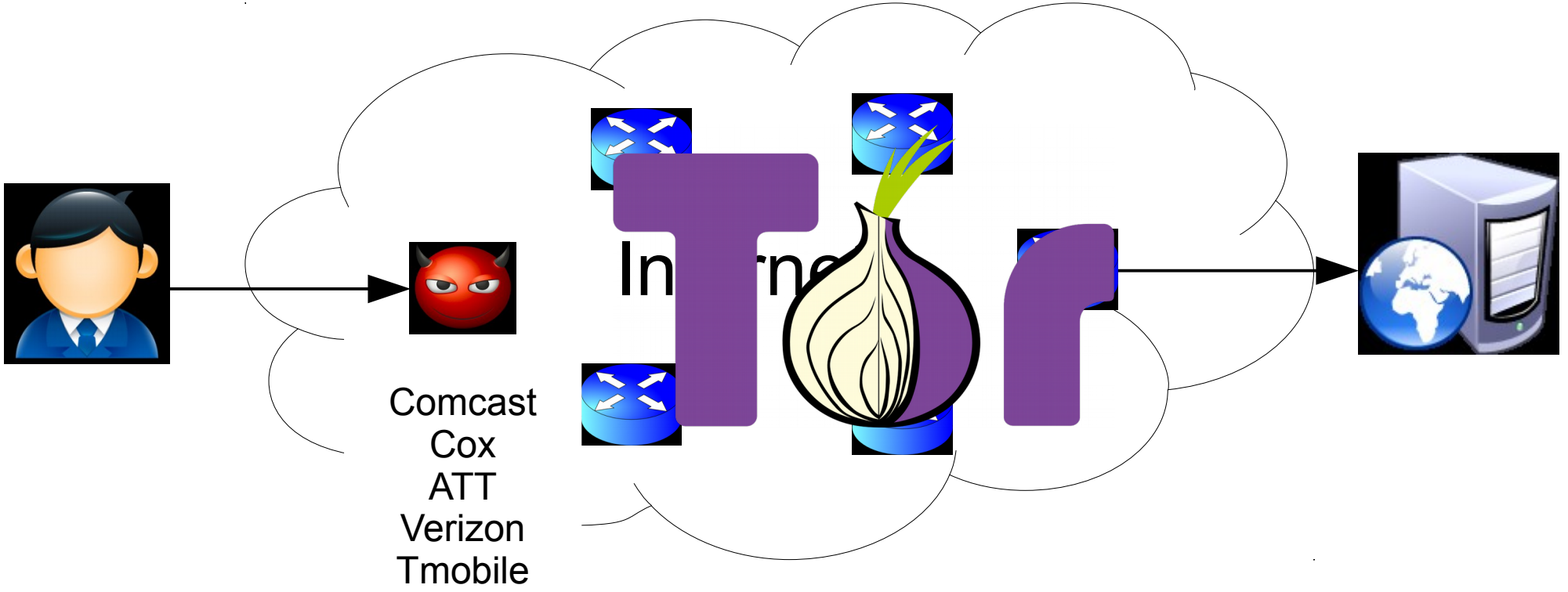
Realistic Cover Traffic to Mitigate Website Fingerprinting Attacks

Weiqi Cui (Oklahoma State University)
Jiangmin Yu (Oklahoma State University)
Yanmin Gong (Oklahoma State University)
Eric Chan-Tin (Loyola University Chicago[1])

[1]Previously at Oklahoma State University

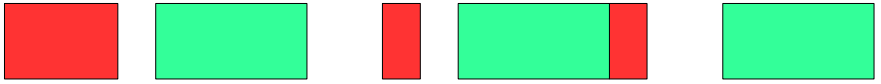

Internet


Internet
VPN
Tor


Comcast
Cox
ATT
Verizon
Tmobile

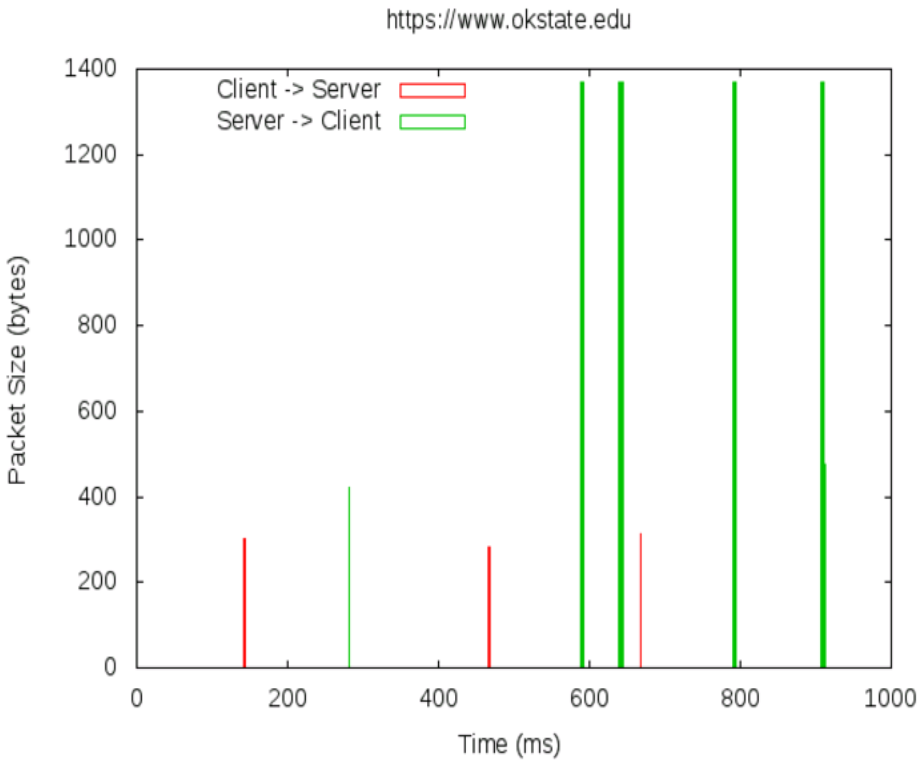
A

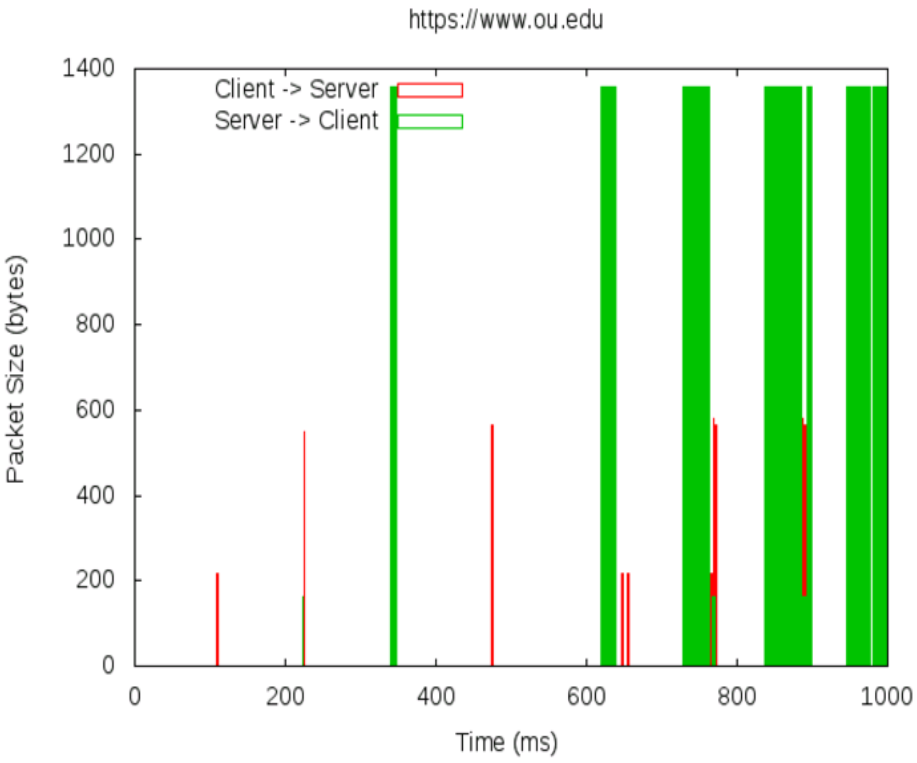
B



C

Website Fingerprinting Attack

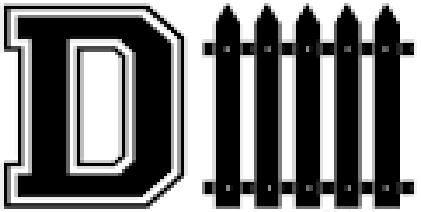




Uses only number of packets, size of packets, and direction of packets

Accuracy

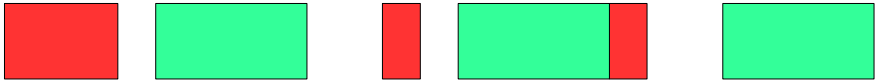
- 80+% accuracy
- Machine learning such as k-NN, SVM, RandomForest



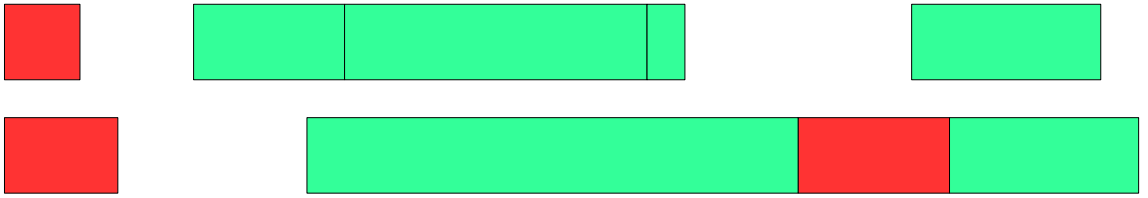
- Padding
  - Every packet has same size
- Delay
  - Same delay
- Extra packets (noise)

A

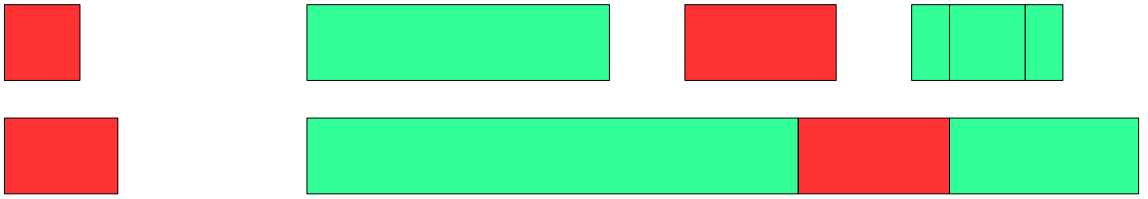
B



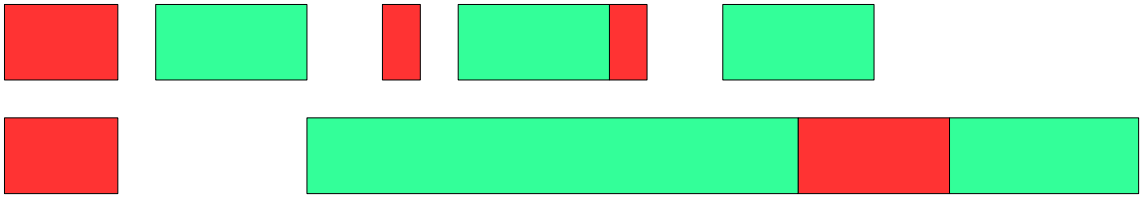
C



A



B

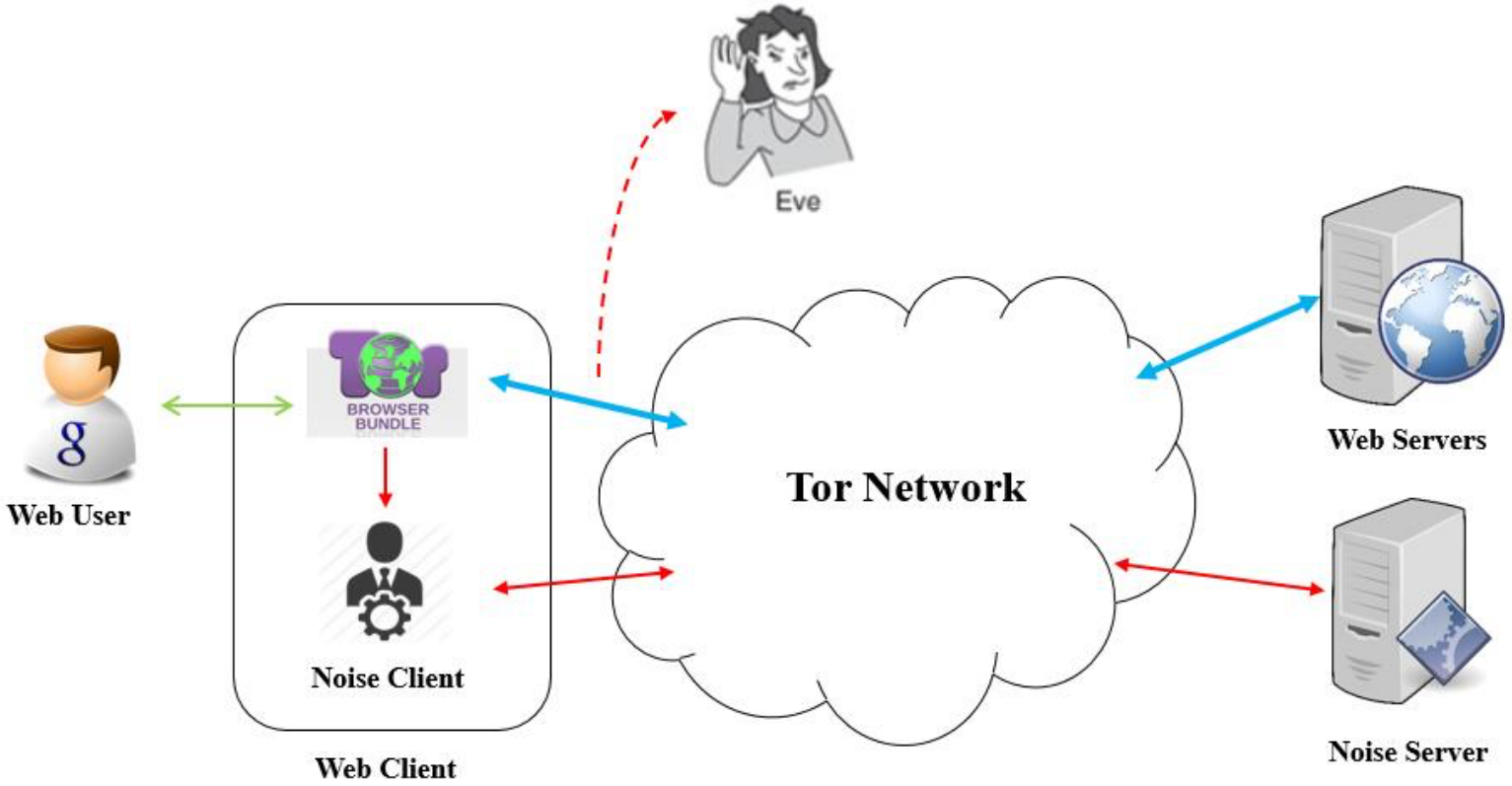


C

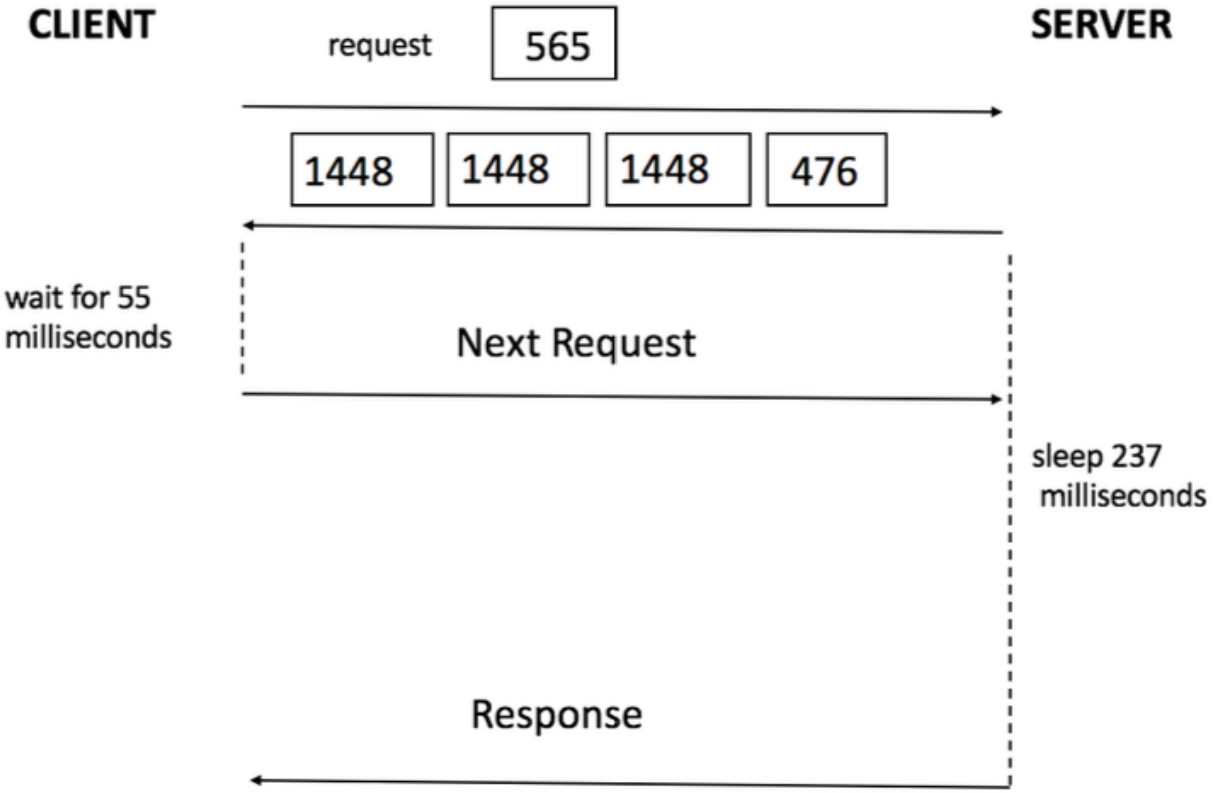
Mitigation	Accuracy	Latency Overhead	Bandwidth Overhead
No defense	91%	0%	0%
CS-BuFLO	22%	173%	130%
Tamaraw	10%	200%	38%
WTF-PAD	15%	0%	54%
Walkie-Talkie	19%	34%	31%

Contribution

- **Motivation**: visiting two websites at once mitigates website fingerprinting attacks
- Use variable realistic network traffic as noise


Eve
BROWSER BUNDLE
Web User
Noise Client
Web Client
Tor Network
Web Servers
Noise Server

83:565	116:565	516:565	4904:565	4905:1448
4905:1448	4907:1448	4907:1448	4908:1448	4931:705
4956:565	4981:1130	5017:565	5018:1448	5018:1448
5019:1448	5020:1448	5022:1448	5022:1448	5024:1448
5024:1448	5025:565	5042:1448	5044:651	5073:1448
5073:1448	5074:1448	5075:1448	5075:1448	5075:565
5079:1448	5079:1448	5098:422	5130:1448	5130:1130
5130:1448	5133:1448	5155:476	5367:565	5388:1448
5388:1448	5391:1448	5395:988	5479:565	5505:1130



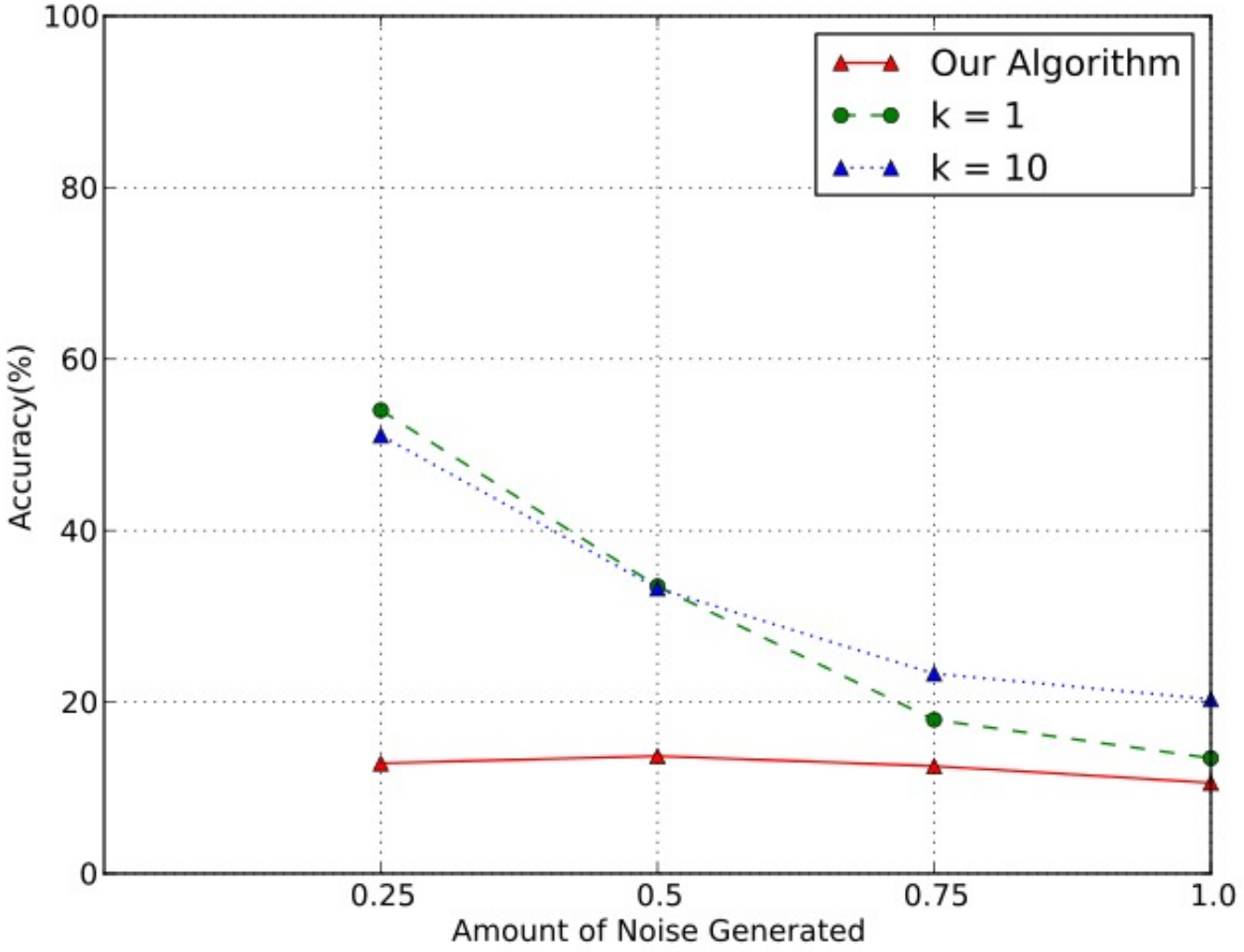
Dataset

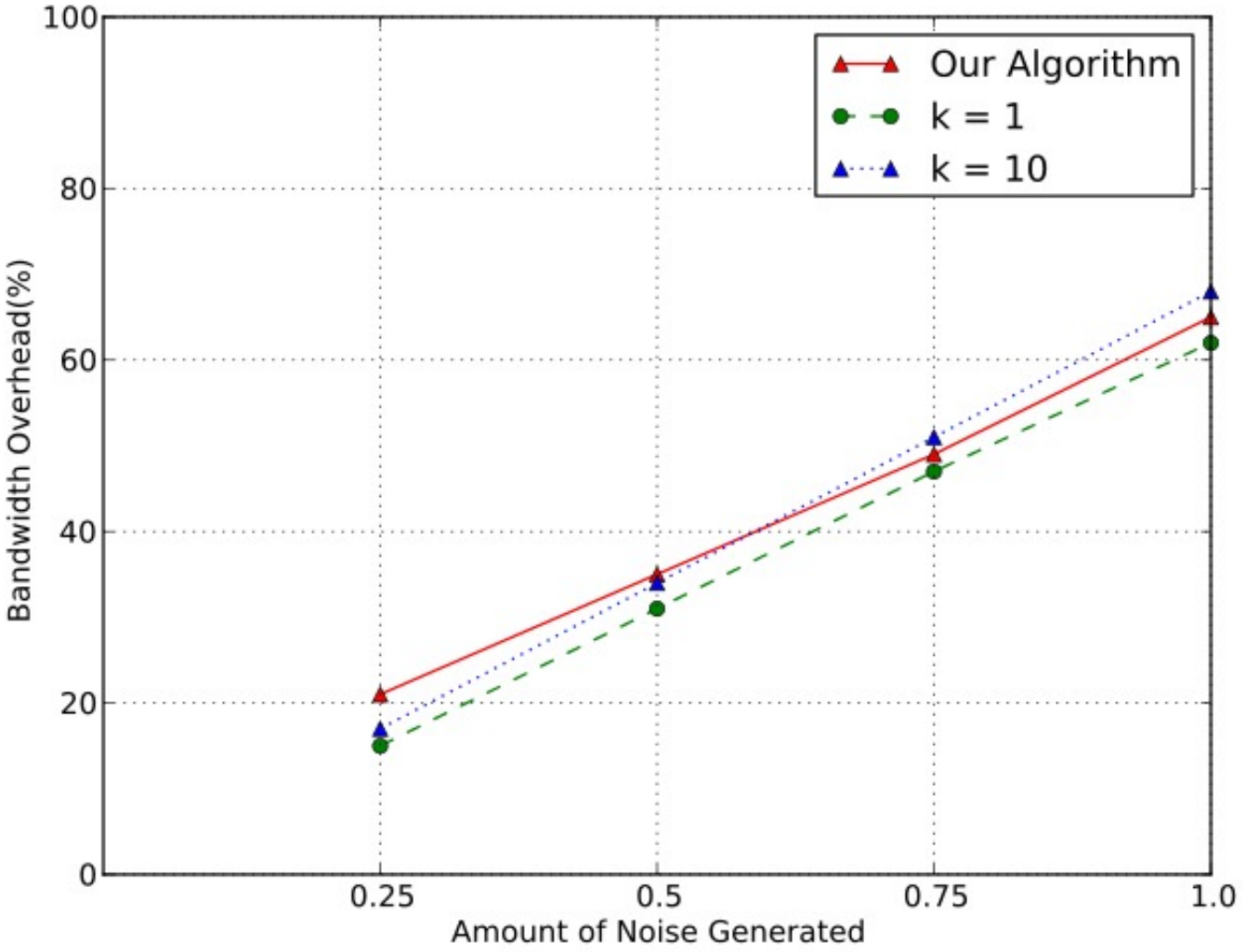
- Panchenko et. al., “Website fingerprinting at Internet scale,” NDSS 2016
- 1,125 websites
  - 40 instances each

Simulation Setup

- 91 websites used as data for cover traffic
- Each simulation run 5 times

Experiments




Our Algorithm
k = 1
k = 10
Bandwidth Overhead(%)
Amount of Noise Generated
0
20
40
60
80
100
0.25
0.5
0.75
1.0

Mitigation	Accuracy	Latency Overhead	Bandwidth Overhead
No defense	91%	0%	0%
CS-BuFLO	22%	173%	130%
Tamaraw	10%	200%	38%
WTF-PAD	15%	0%	54%
Walkie-Talkie	19%	34%	31%
Our Algorithm	14%	0%	20%

Discussion

- Noise server used instead of real webserver
- Noise is empty packets – only packet size and number of packets matter
- Could be based on real user traffic or on a record of K generic websites

Summary

- Realistic cover traffic
- Mitigates website fingerprinting attacks
- Low bandwidth overhead
- No latency overhead
- Run real-world experiments instead of simulations

Acknowledgments

Thank you!

Questions?

Eric Chan-Tin, chantin@cs.luc.edu