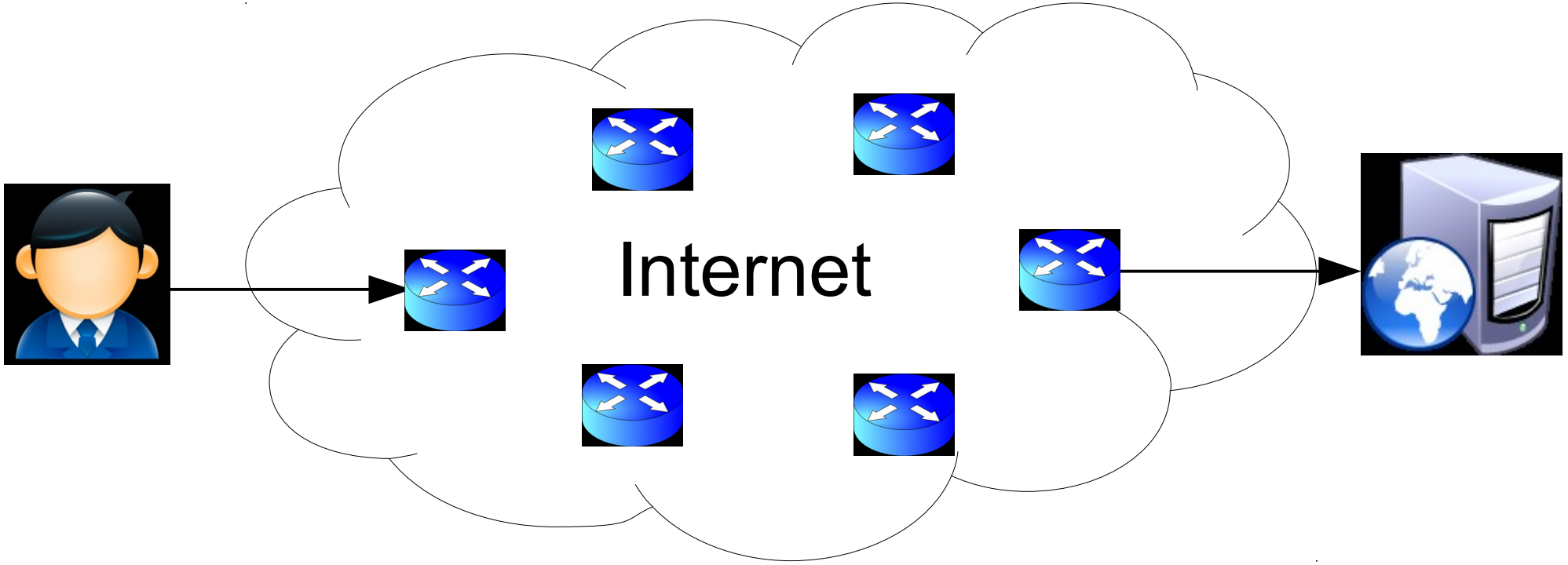
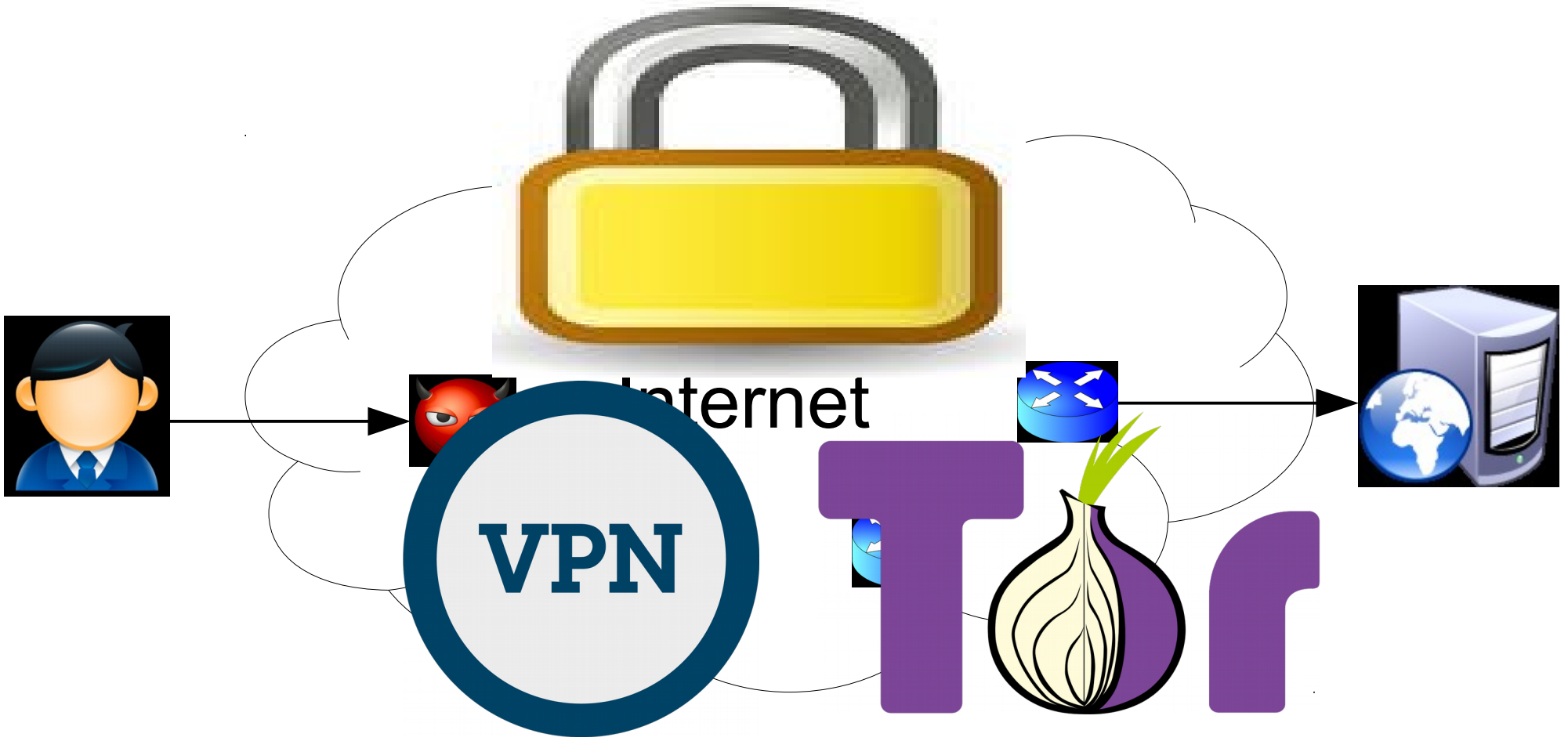
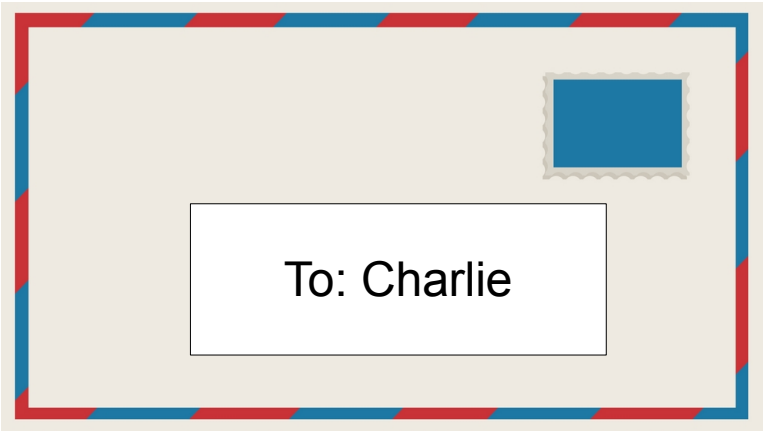
Website Fingerprinting Attack Mitigation using Traffic Morphing

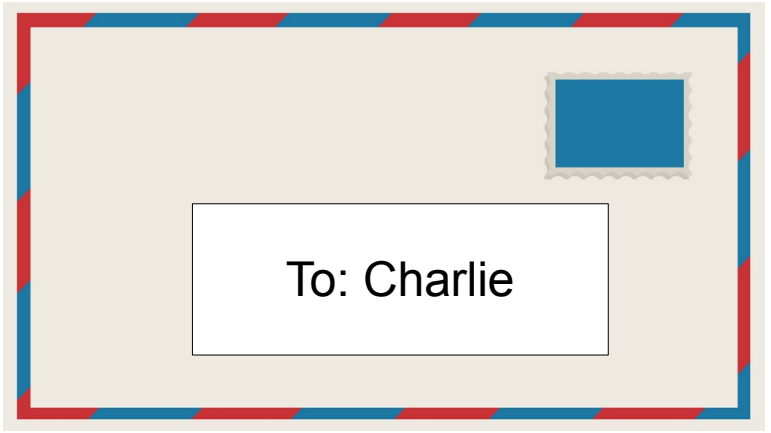
Eric Chan-Tin (Loyola University Chicago[1])
Taejoon Kim (Texas A&M University, Commerce)
Jinoh Kim (Texas A&M University, Commerce)

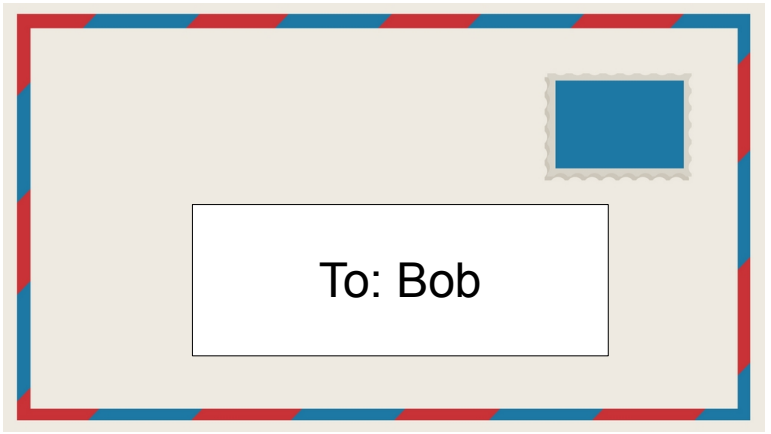
[1]Previously at Oklahoma State University

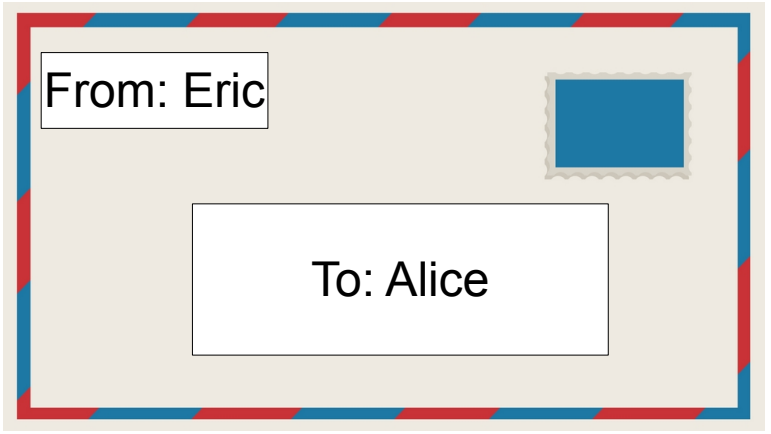

Internet

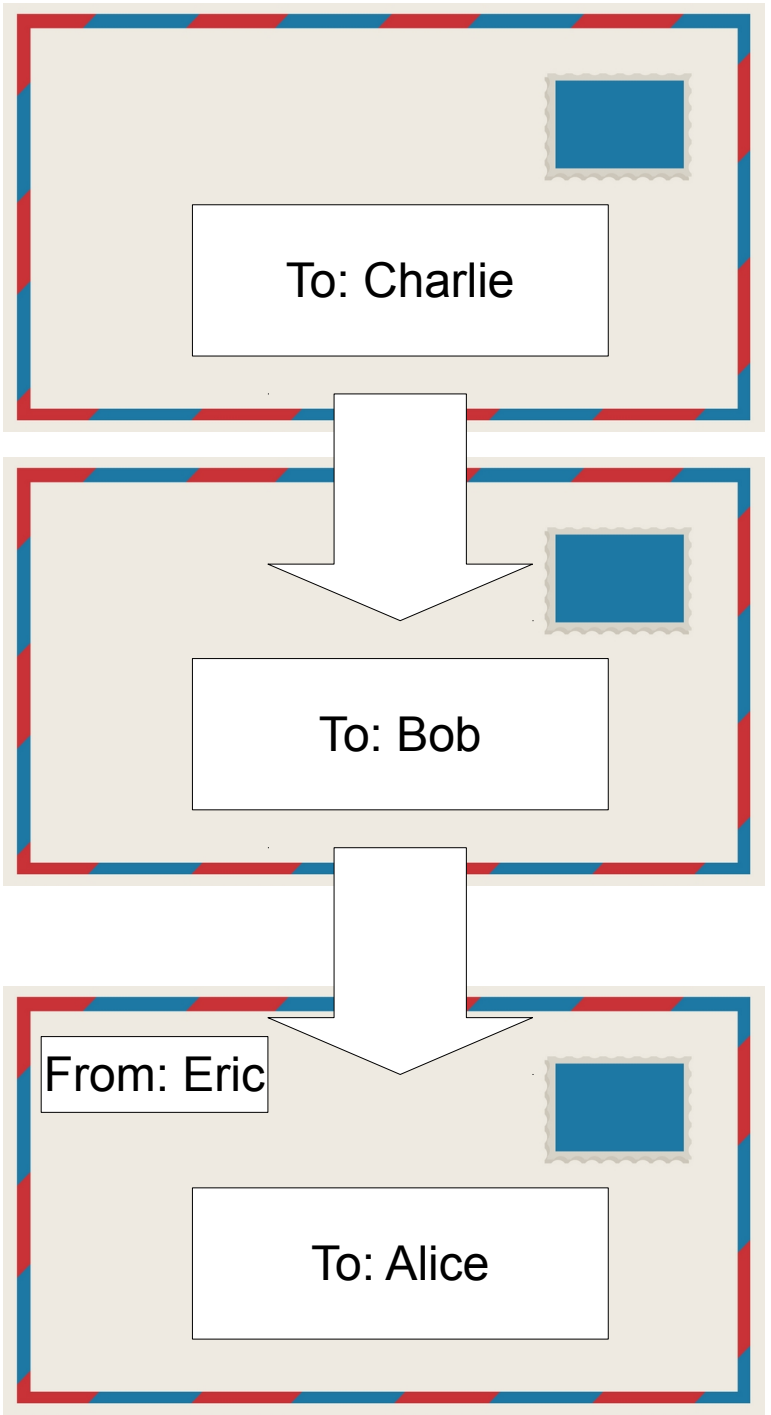

Internet
VPN
Tor

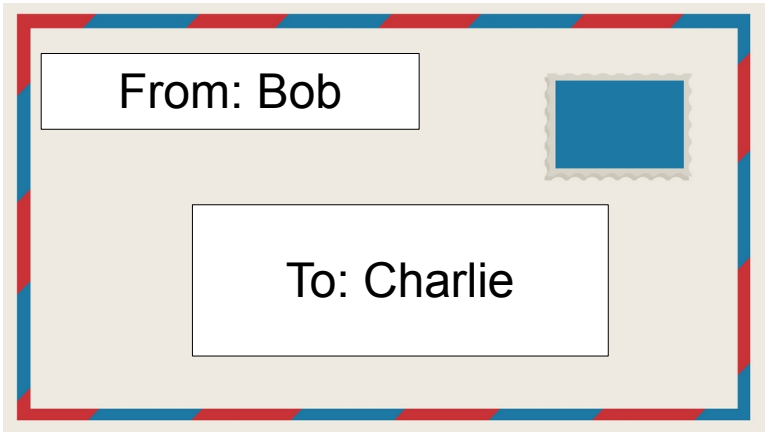

To: Charlie

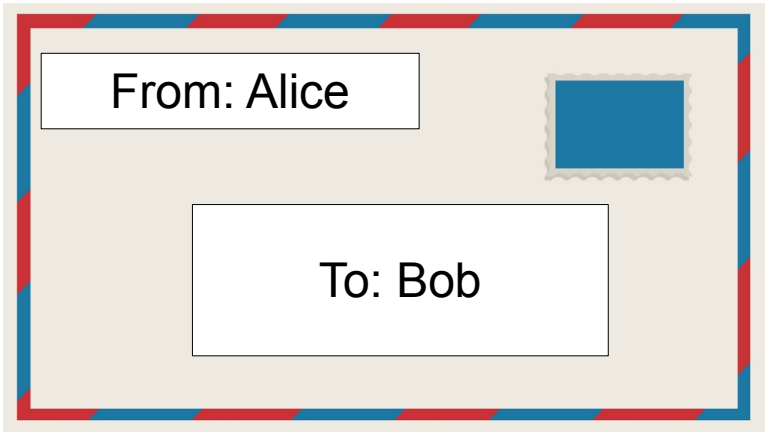

To: Charlie

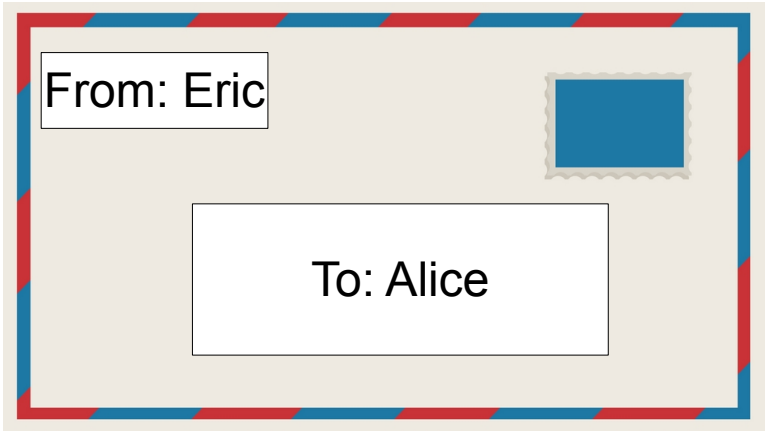

To: Bob


From: Eric
To: Alice


To: Charlie
To: Bob
From: Eric
To: Alice

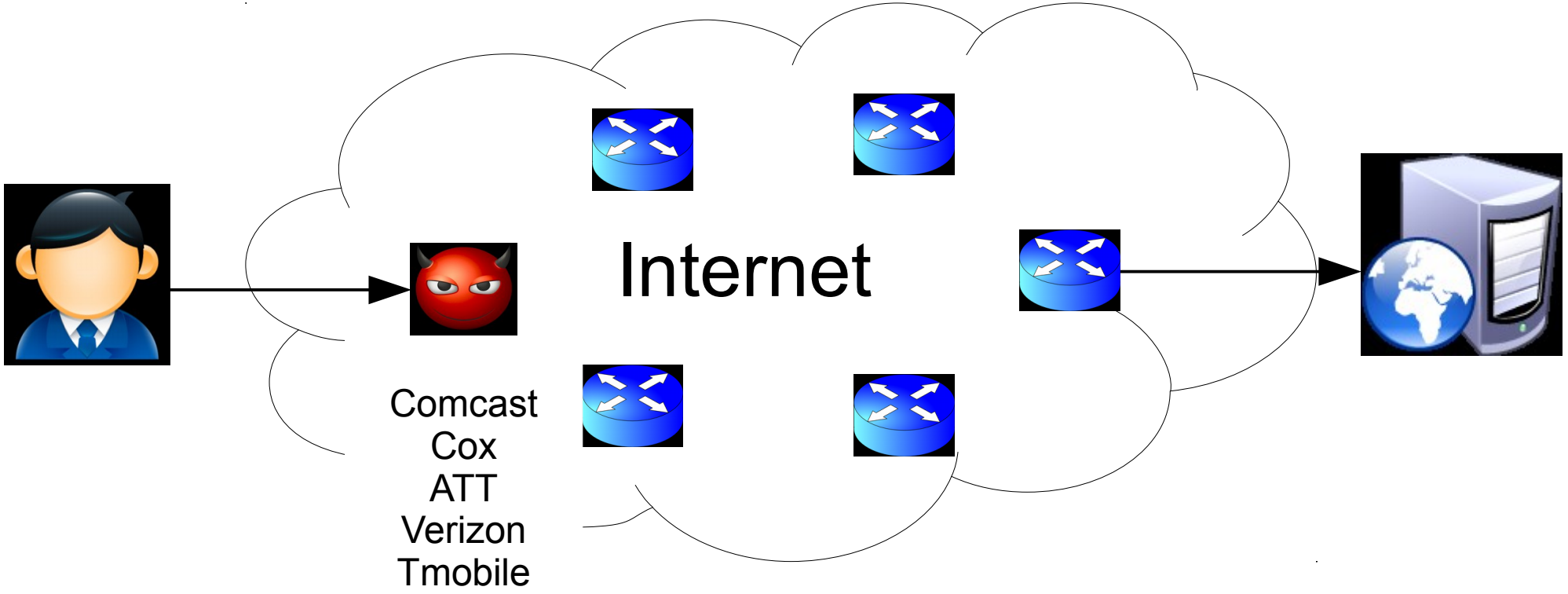

From: Bob
To: Charlie


From: Alice
To: Bob


From: Eric
To: Alice

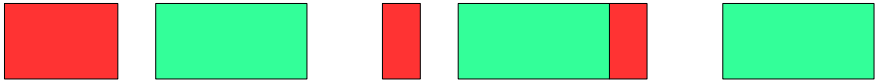
Tor

- Free, open-source
- 2002
- Anonymity network
- Onion routing
- 2,000,000+ users daily
- 7,000+ volunteers (relay nodes)


Internet
Comcast
Cox
ATT
Verizon
Tmobile

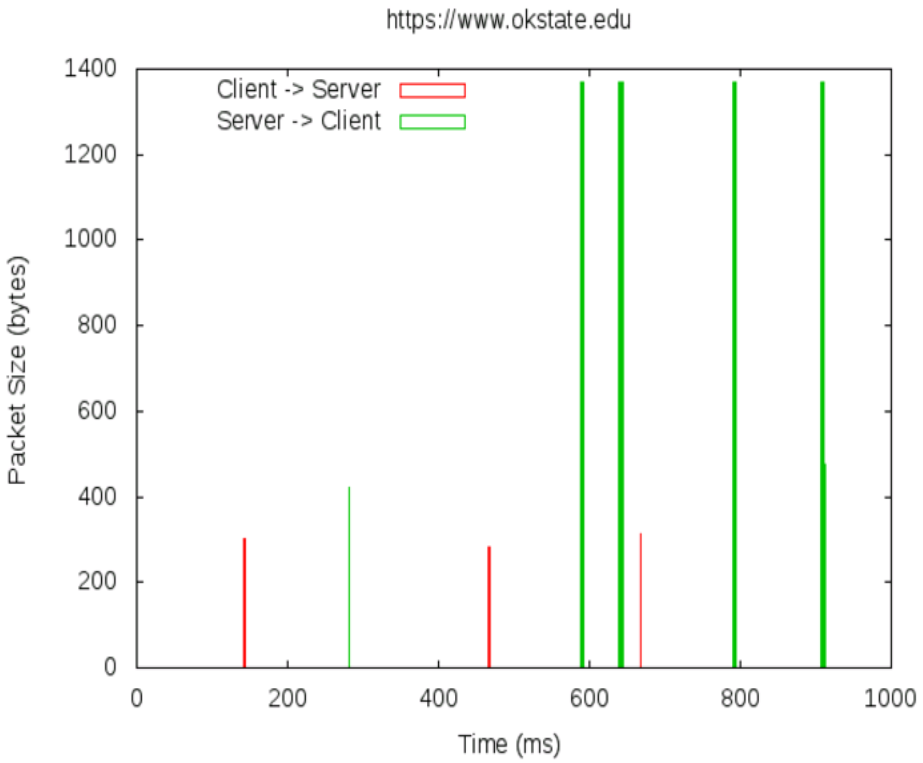
A

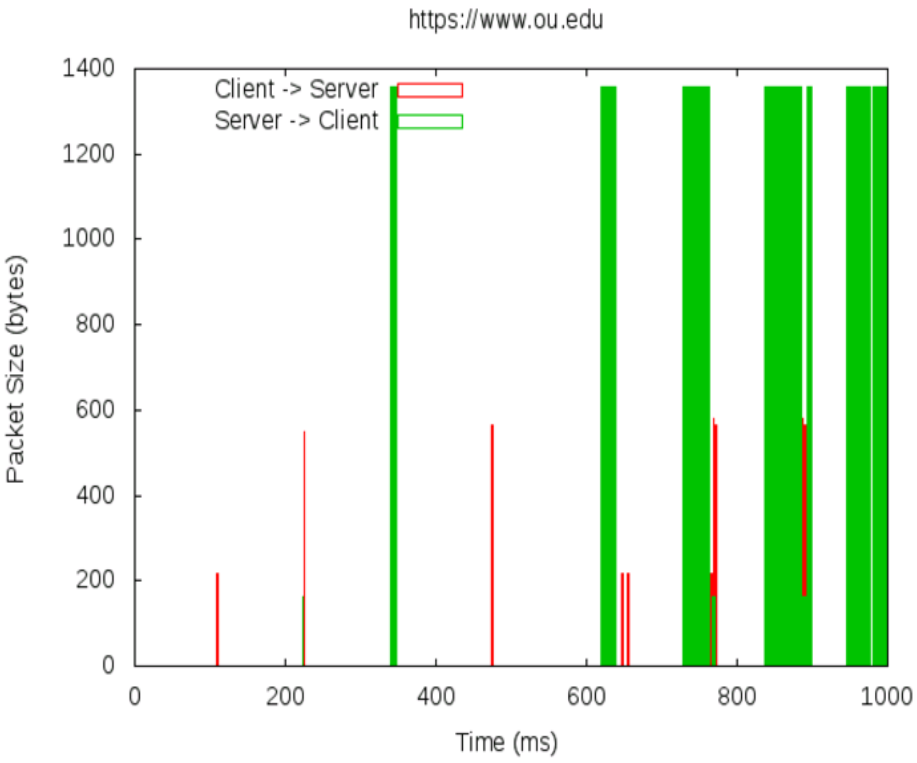
B



C

Website Fingerprinting Attack

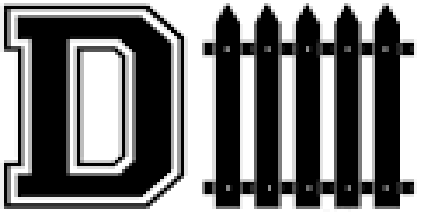




Uses only number of packets, size of packets, and direction of packets

Accuracy

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



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

- Make every website look similar using traffic morphing

But...

- Hard to make every website on the Internet look similar

Contribution

- Some websites already look "similar"
  - Number of packets
  - Size of packets
- Cluster "similar" websites and make all websites within a cluster indistinguishable
  - Easier to do within a cluster than for ALL websites on the Internet
  - Use traffic morphing

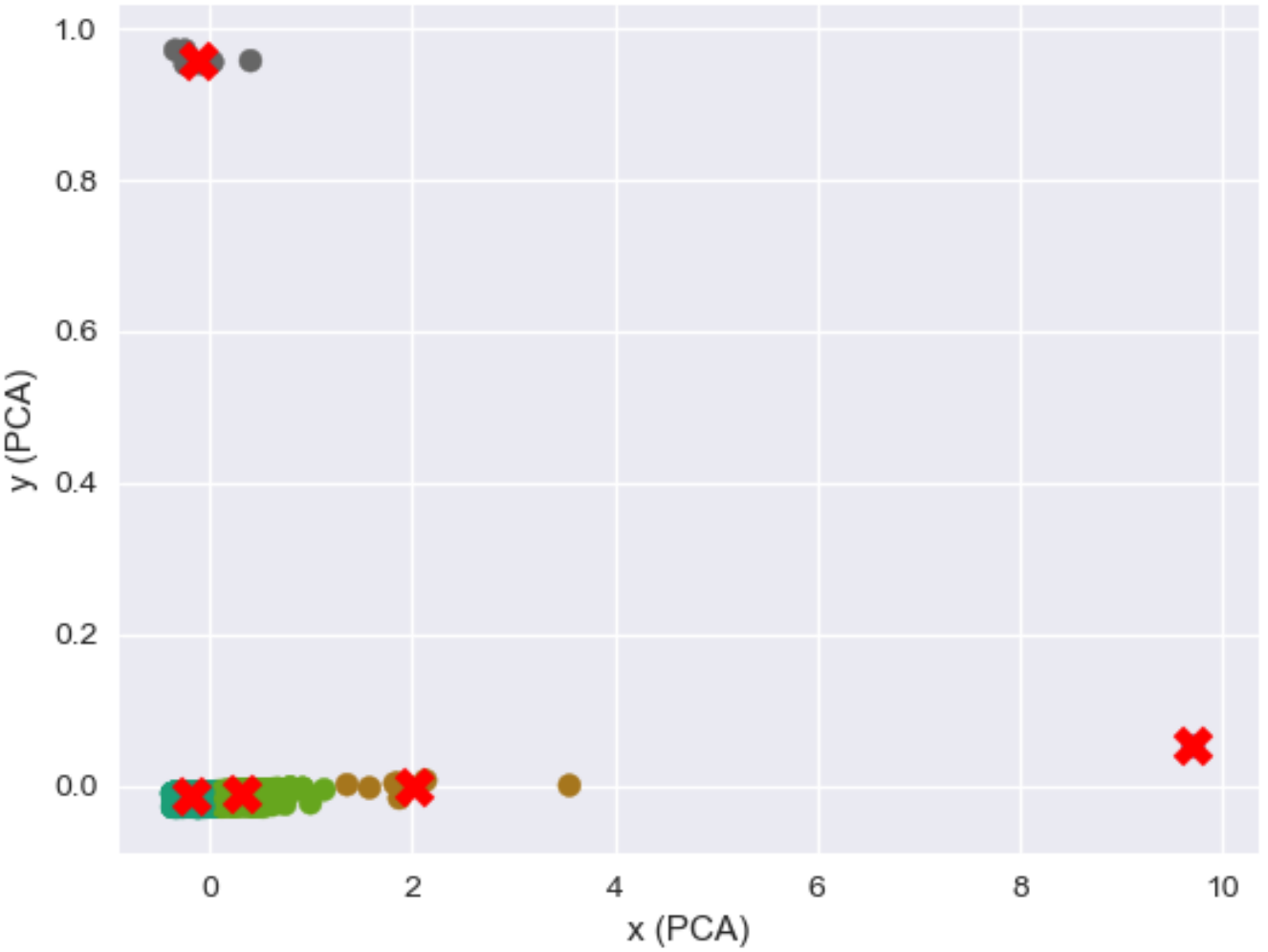
Dataset

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

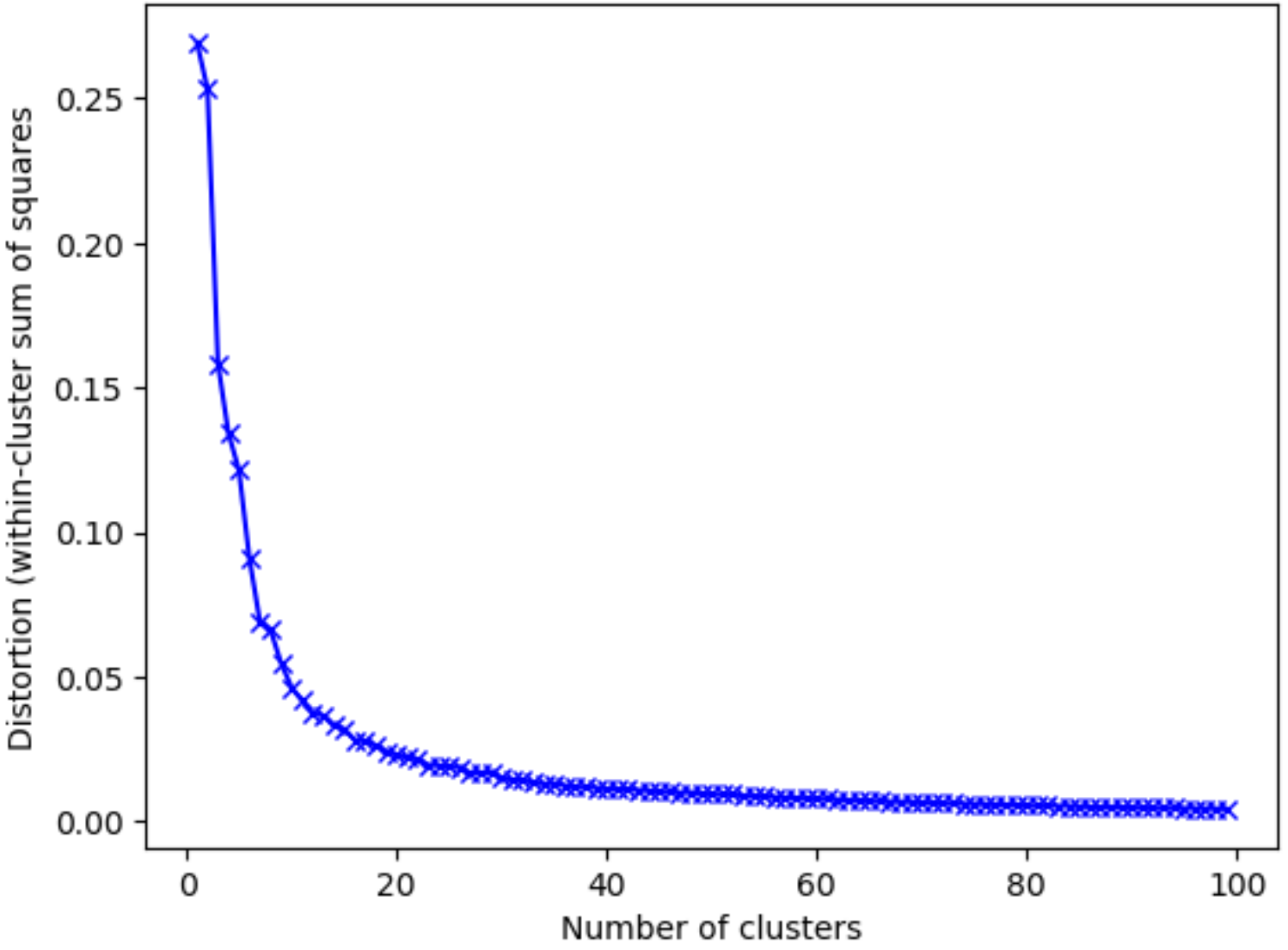
Proposed Algorithm

- Cluster websites
  - PCA (Principal Component Analysis)
  - 104 features
    - Total number of outgoing packets
    - Total number of incoming packets
    - Total size of all outgoing packets
    - Total size of all incoming packets
    - 100 samples of cumulative packet sizes
- Traffic morph each cluster to make all websites within that cluster indistinguishable

Clustering



Elbow Method



Traffic Morph Method

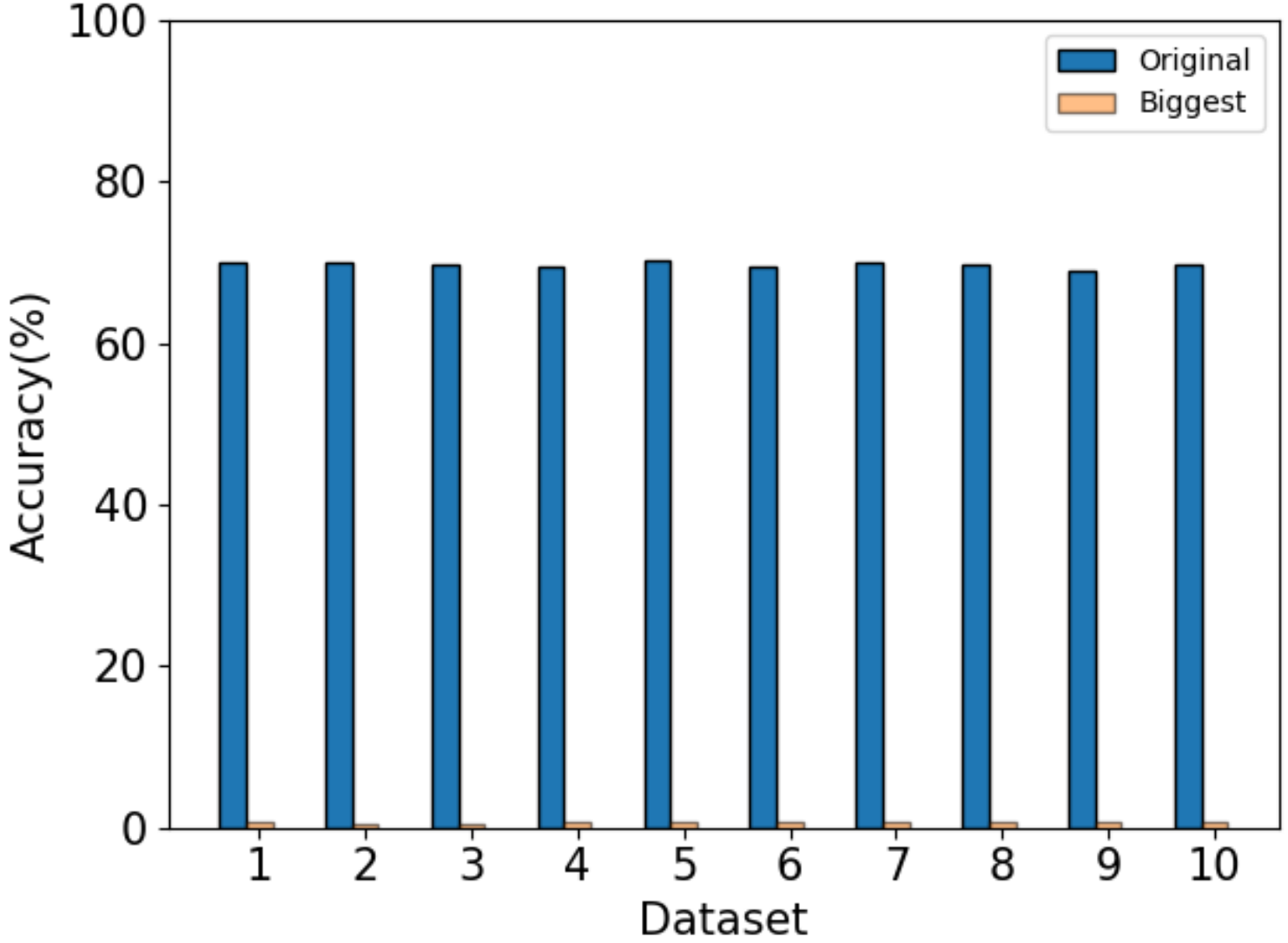
- *Biggest*
- Make the size of every packet in a cluster be the same size as the biggest packet in that cluster

Experiment Setup

- 50% of dataset for training, 50% for testing
- Repeat 10 times

Accuracy

Mitigation	Accuracy
No defense (Tor)	91%
CS-BuFLO	22%
Tamaraw	10%
WTF-PAD	15%
Walkie-Talkie	19%
Our Algorithm	< 1%


Original
Biggest
100
80
60
40
20
0
Accuracy(%)
1
2
3
4
5
6
7
8
9
10
Dataset

Overhead

Mitigation	Latency Overhead	Bandwidth Overhead
No defense (Tor)	0%	0%
CS-BuFLO	173%	130%
Tamaraw	200%	38%
WTF-PAD	0%	54%
Walkie-Talkie	34%	31%
Our Algorithm	0%	210%

Summary

- Promising research
  - Clustering
  - Traffic morphing within each cluster
- Almost completely mitigates website fingerprinting attacks (<1% accuracy)
- Other methods such as average packet size, random packet size, and closest packet size

Thank you!

Questions?

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