Website Traffic Analysis

Michael Freyberger and Michael Chang

ELE 298 Advisor: Professor Prateek Mittal

Princeton University | Electrical Engineering | Class of 2016

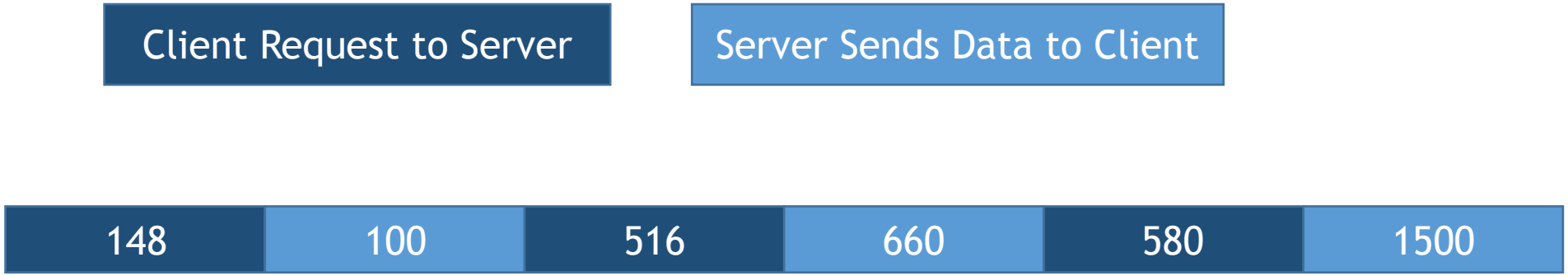
Motivation

- Anonymity networks are becoming increasingly important
  - Free network access in oppressed countries is desirable
  - Prevent marketers from tracking web activity
- Current anonymity networks are vulnerable

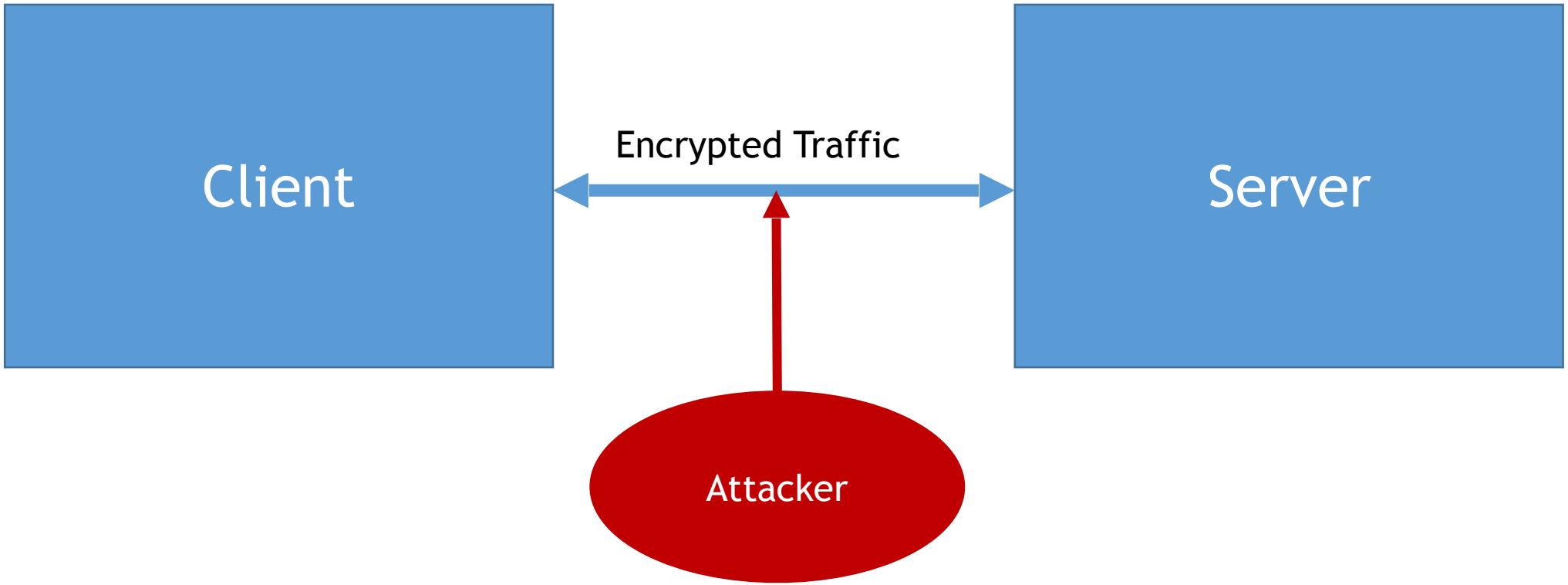
Basic Communication

Characteristics of Communication

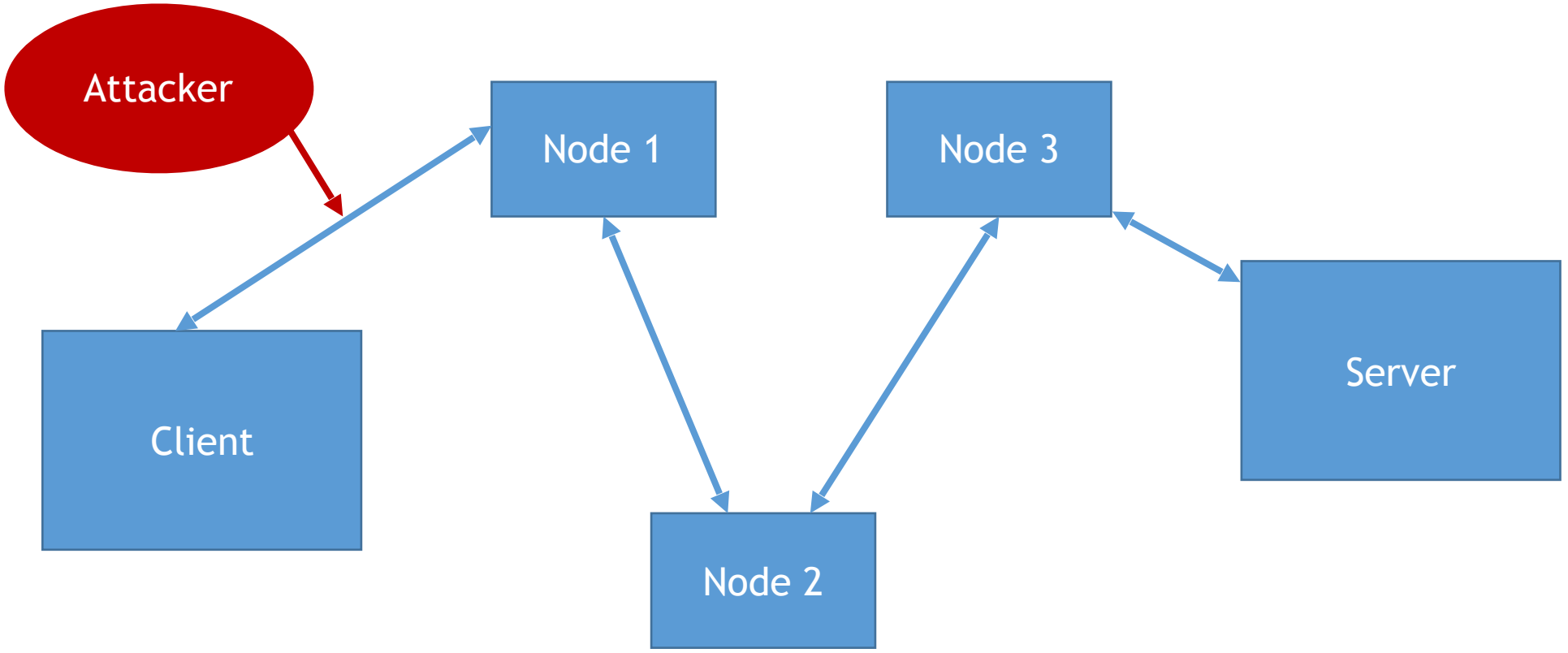
- Information can be readily encrypted
- Data transfer is broken down into packets
  - Packets have a direction and length
  - Packets have a maximum length called the Maximum Transmission Unit (MTU)
    - A typical size for the MTU is 1500 bytes



Scenario 1: Possible Attack

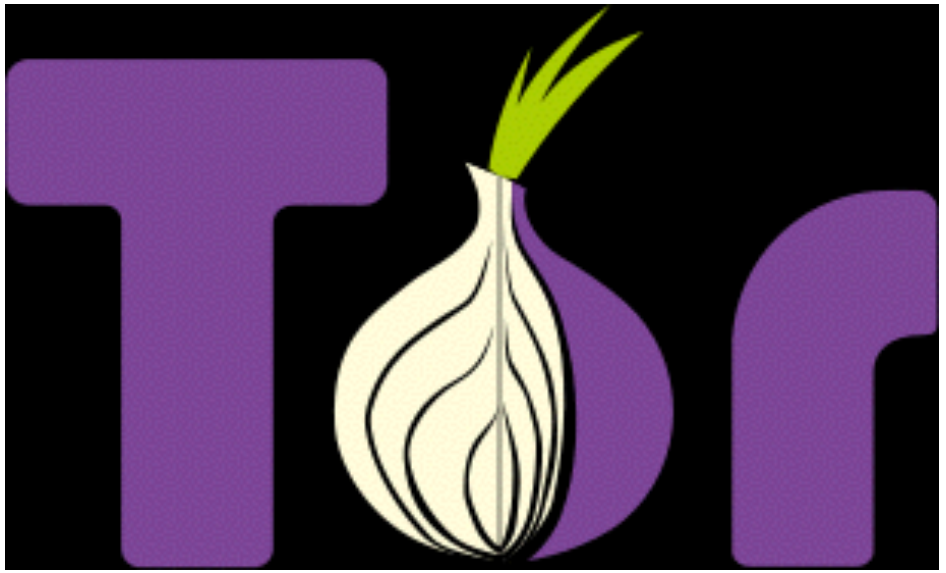


Scenario 2: Possible Attack with Proxies



Current Secure Systems

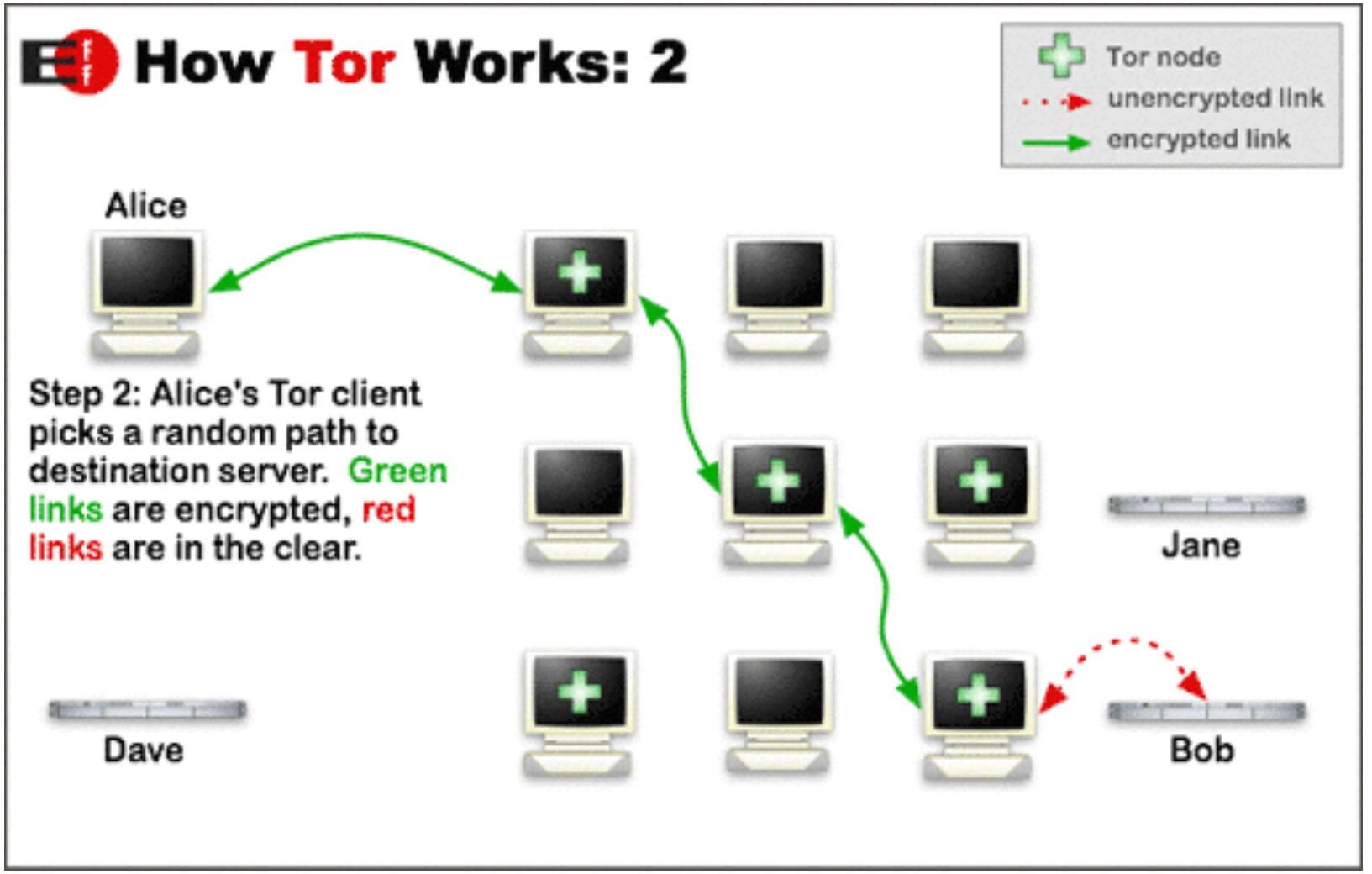
- Cryptographic Security
  - SSL
  - HTTPS
- Onion Routing Systems
  - Tor
  - JonDonym



www.torproject.org

Tor Project

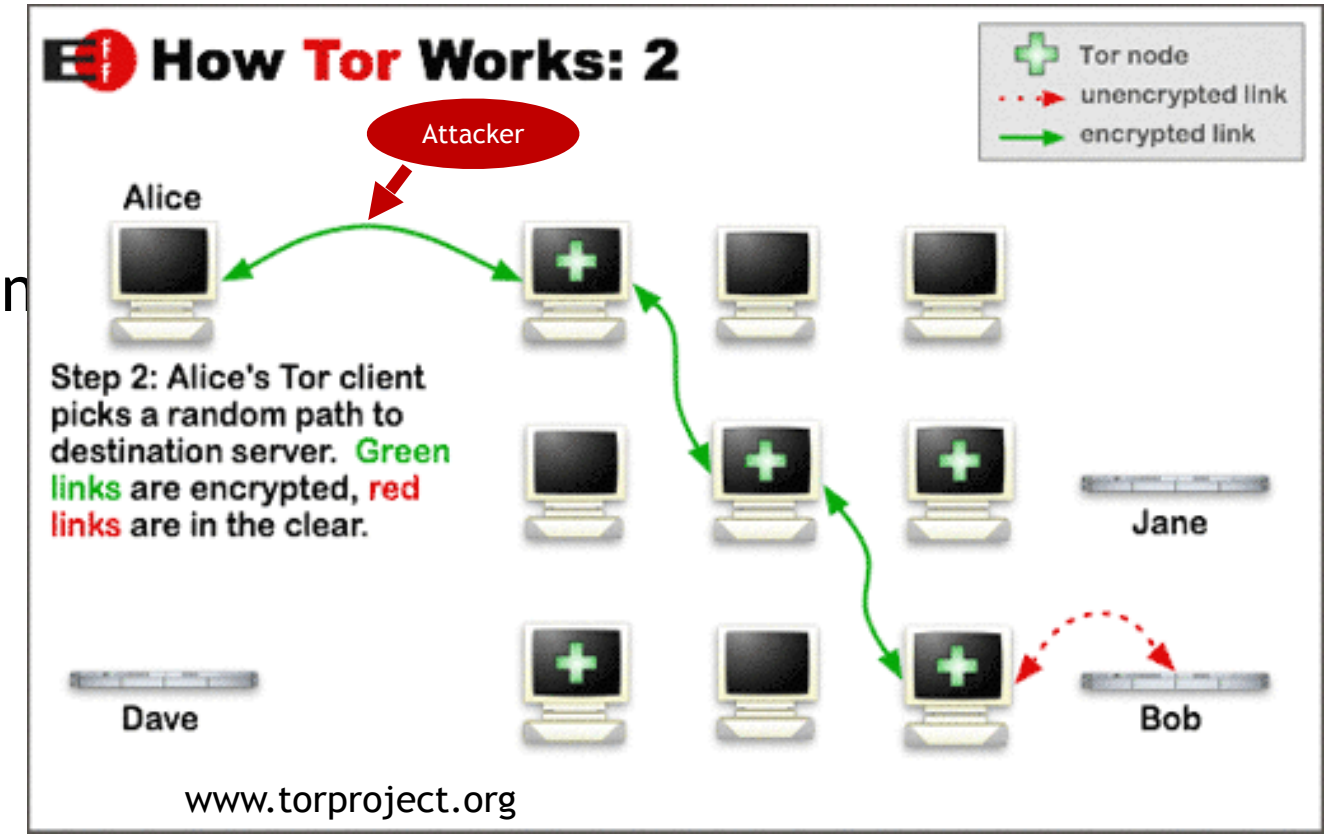
- 2 Million Users
- Pads all data cell units to 512 bytes
- Randomly generates 3 proxies



www.torproject.org

Traffic Analysis Attacks

- Single Attacker
- Exploit metadata of the communication
  - Total Number of Bytes
  - Packet Lengths and Direction
  - Packet Bursts



Panchenko et al. [1]

- Combined packet and burst level approach
- Support Vector Machine
  - Many features utilized, but primary ones:
    - Total Transmitted Bytes
    - Number of packets per burst
    - Count of each packet size
    - Percentage incoming packets
- Achieved accuracy of 47.36% on closed world data set

Cai et al. [2]

- Damareau-Levenshtein Edit Distances
  - Considers natural re-ordering of the packets during information exchange
  - Features considered
    - Packet Sizes
    - Direction of Packets
- Classification using SVM (support vector machines)

Dyer et al.[3]

- Coarse features contain a lot of information
  - Total Time
  - Total Bandwidth
  - Bursts
- Coarse features are difficult to hide
  - BufLo
  - High Overhead

Recreating Tests and Running our own

- Recreating Cai's Classifier
  - Potential Improvements
    - Damereau-Levenshtein Distance transformation penalties
    - Burst-level calculations
- Recreating Dyer's Data set
  - Successful
  - Data Set Setting retrieved from Tor Nodes
    - 8510 websites collected at a rate of 85/hour
    - Linux client machine
    - Browser: Firefox 2.0

Accuracy and Overhead Calculations

- $Accuracy = {}^{c}/_{Tk}$
  - C: the number of correctly classified test traces
  - K: is our privacy set size
  - T: is the number of test traces per site.

- $\text{Overhead} = \dfrac{Bytes\ in\ Pre{-}Countermeasure\ Trace}{Bytes\ in\ Post{-}Countermeasure\ Trace}$

Previously Explored Countermeasures

- Packet Padding Level Padding
  - Pad to MTU
  - Random Padding
  - Exponential Padding
- Sending Dummy Packets
  - BuFLO - Send fixed length packets at fixed time intervals

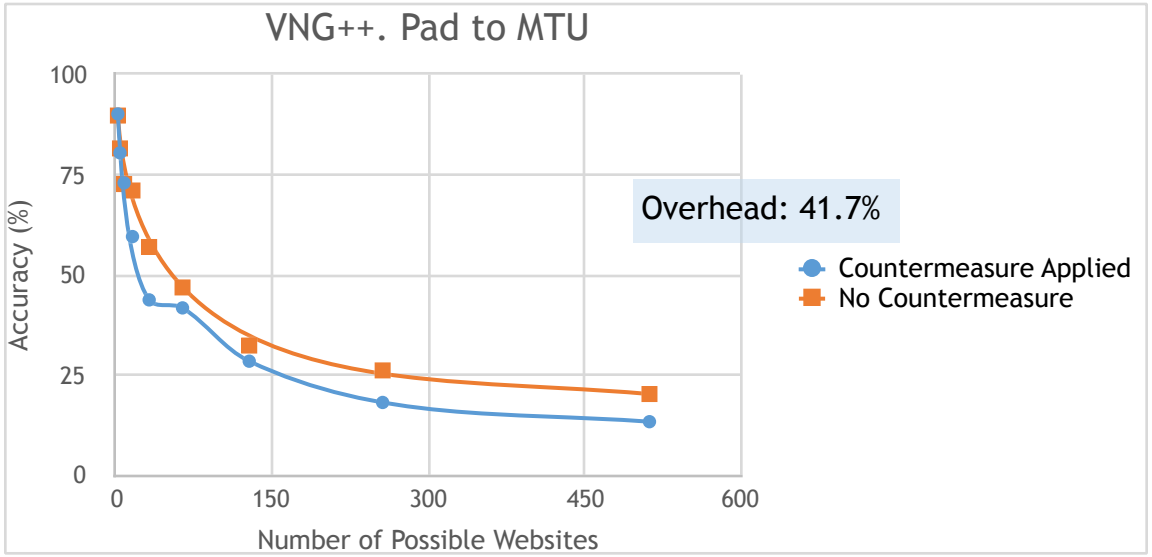
Pad to MTU

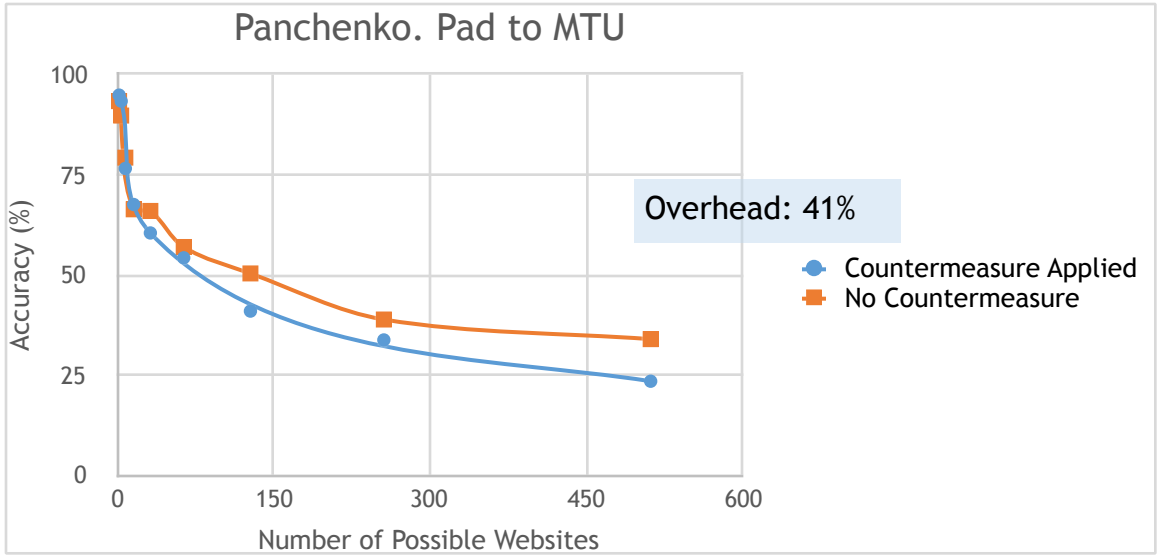
No Countermeasure:

148	100	516

Countermeasure Applied:

1500	1500	1500





Session Random 255

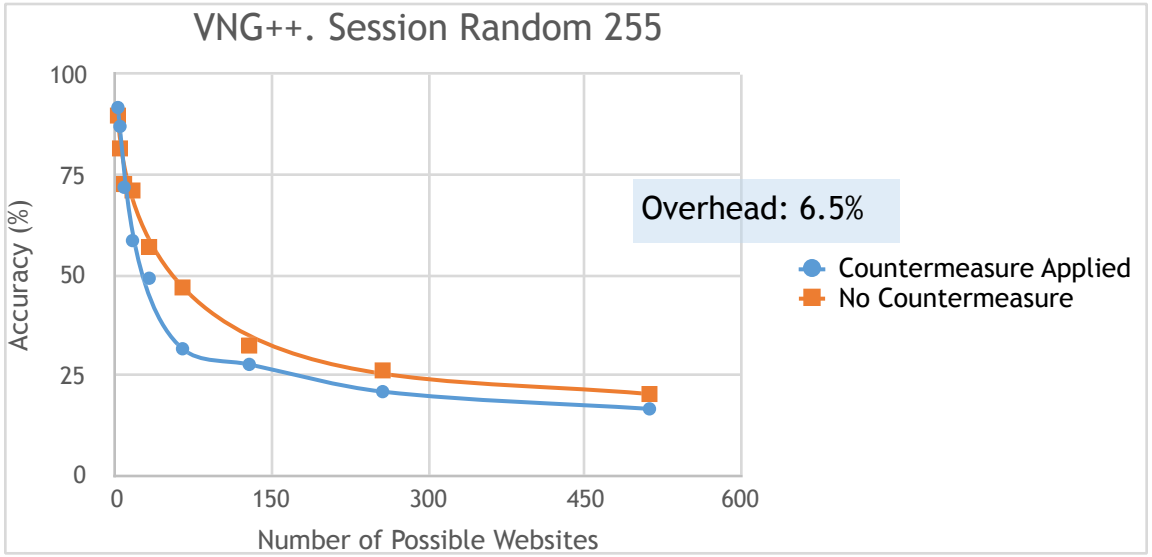
No Countermeasure:

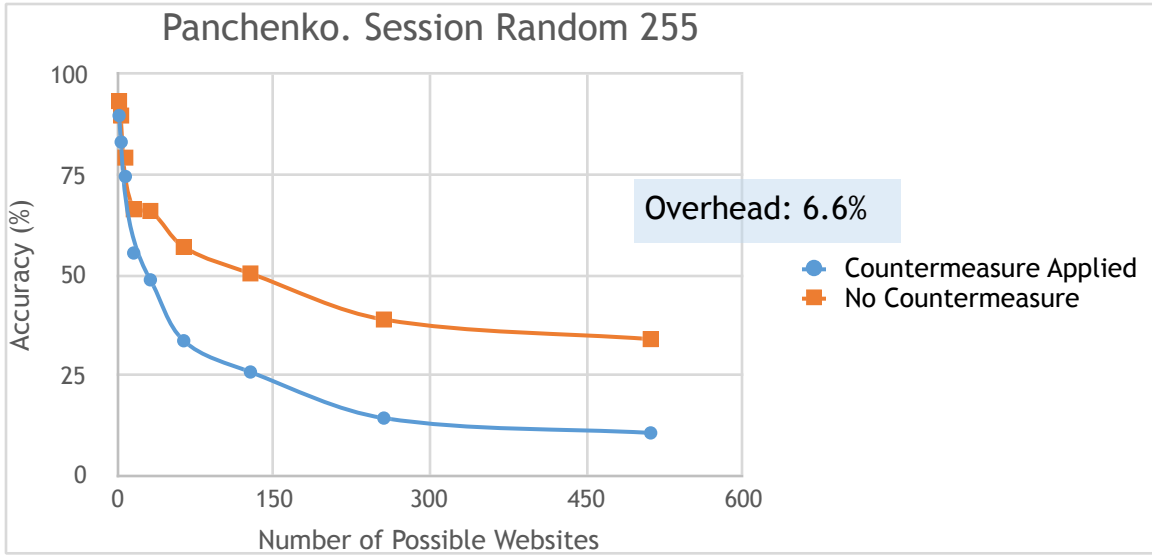
148	100	516

Countermeasure Applied:

236	188	604

Parameters:
P = 88





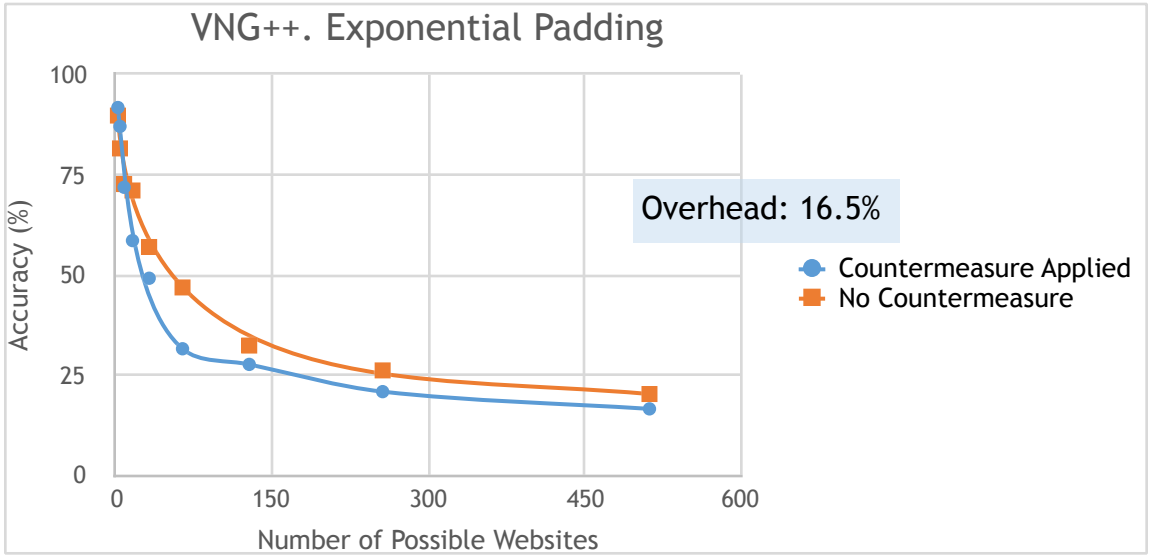
Exponential Padding

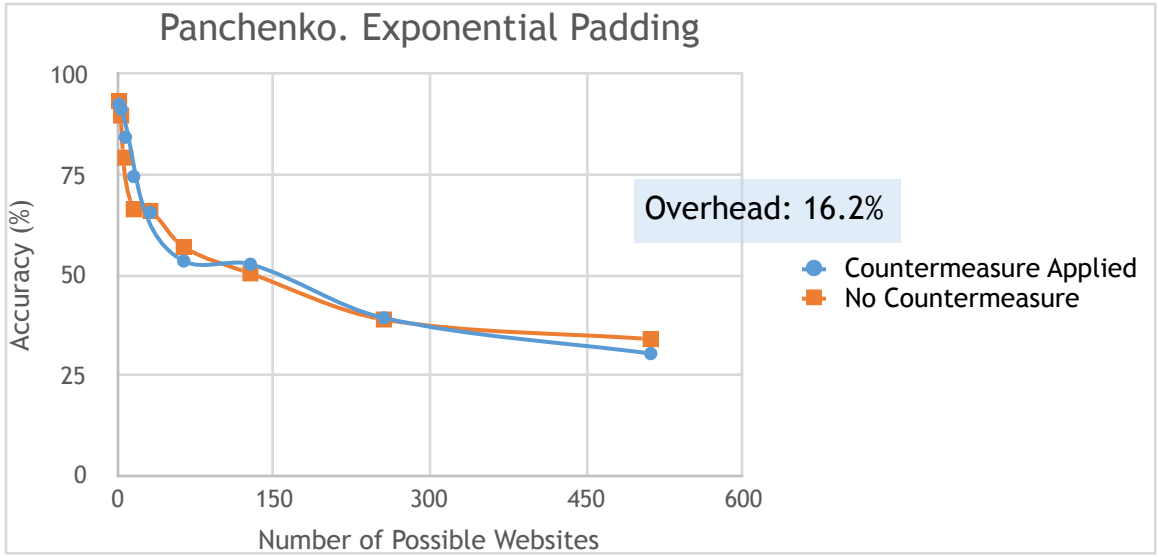
No Countermeasure:

148	100	516

Countermeasure Applied:

256	128	1024





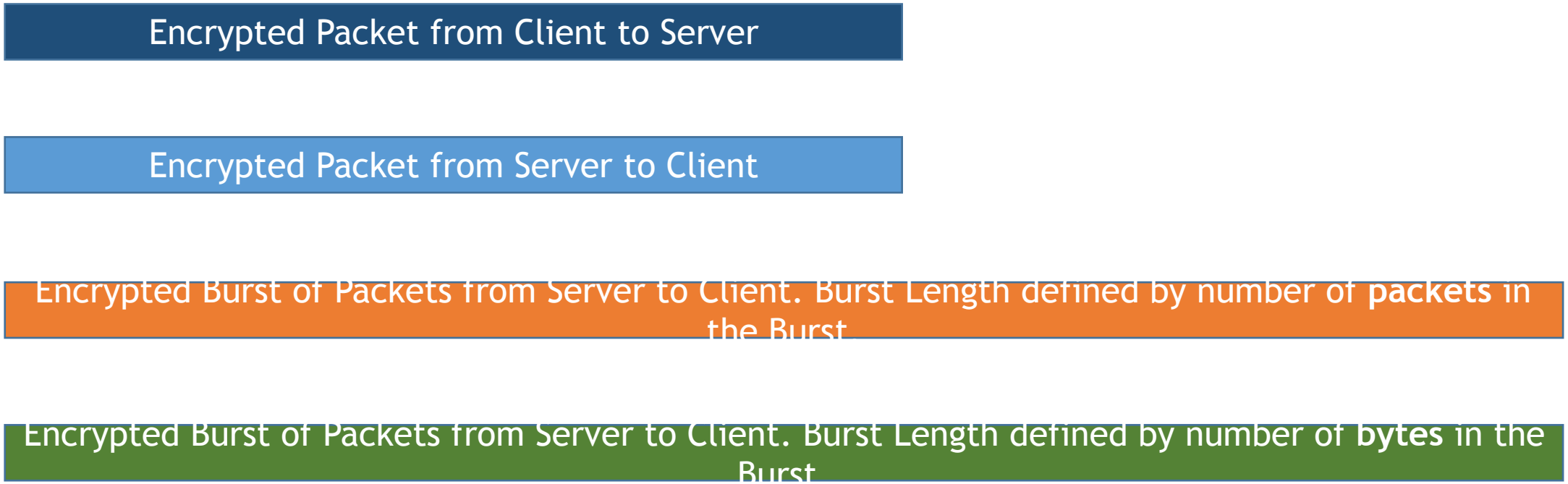
Packet Level Countermeasures

- **Ineffective** – cannot decrease accuracy when classification uses top performing classifiers
- **High Overhead –** Cannot decrease accuracy even with maximal overhead
- **Solution:** Must apply countermeasures at the burst level

Novel Countermeasures

- Burst Level Padding – strategically placed Dummy Packets
- Burst Level Padding in addition to Packet Level Padding
- Random Burst Level Padding
- Bursts defined by number of packets or number of bytes
- Sending Dummy Packets in one direction – Client to Server

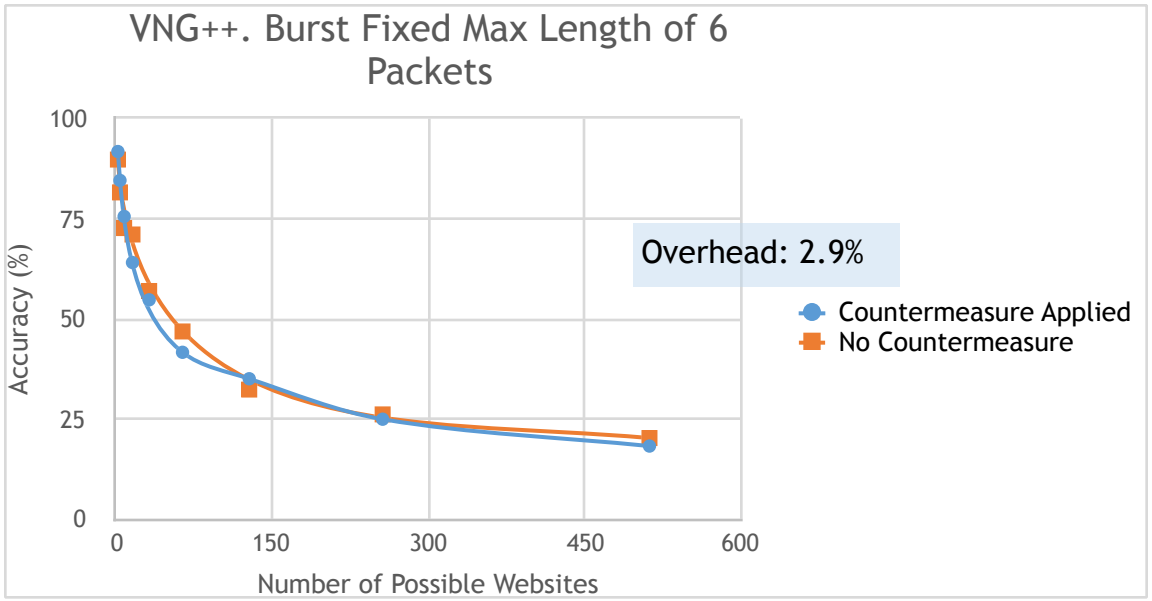
Communication Legend

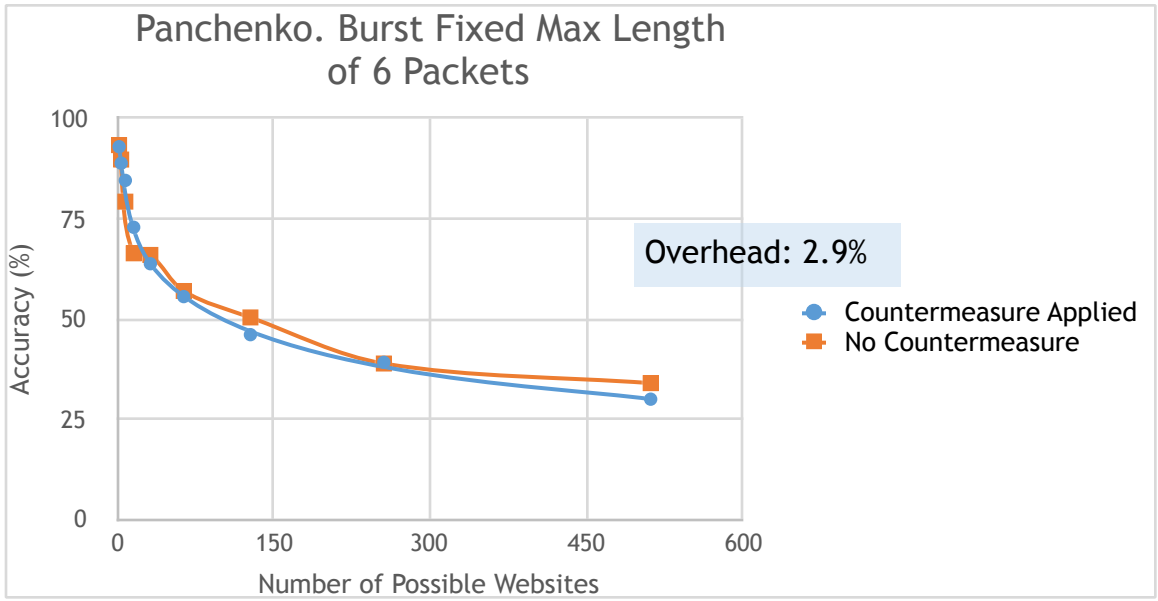


Burst Fixed Max Length of 6 Packets

No Countermeasure:

Countermeasure Applied:



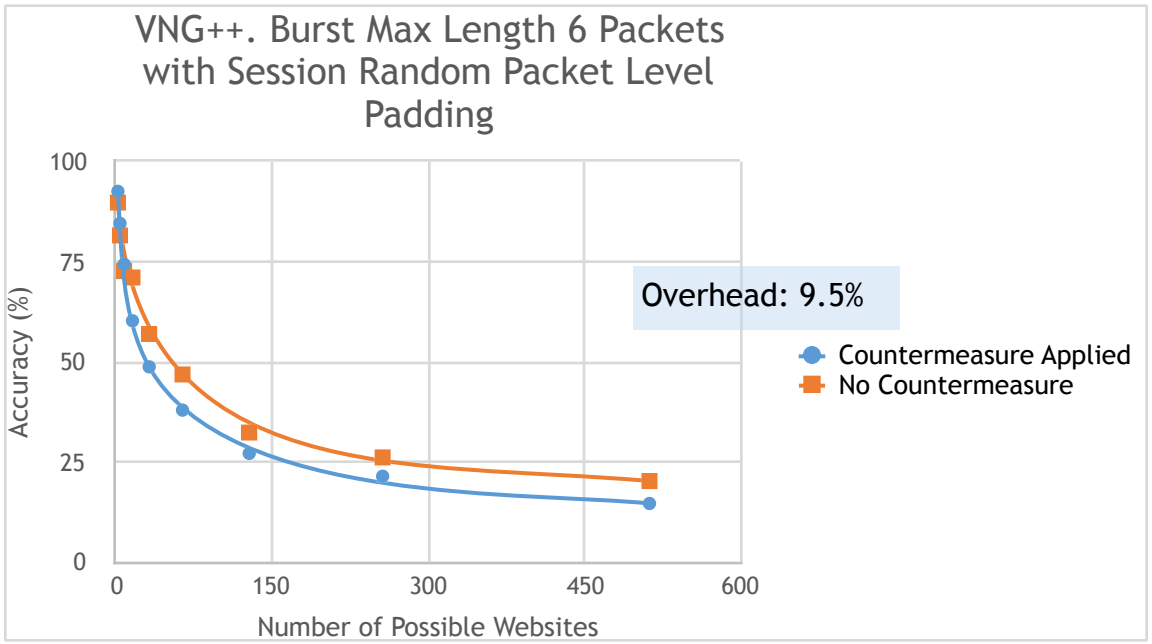


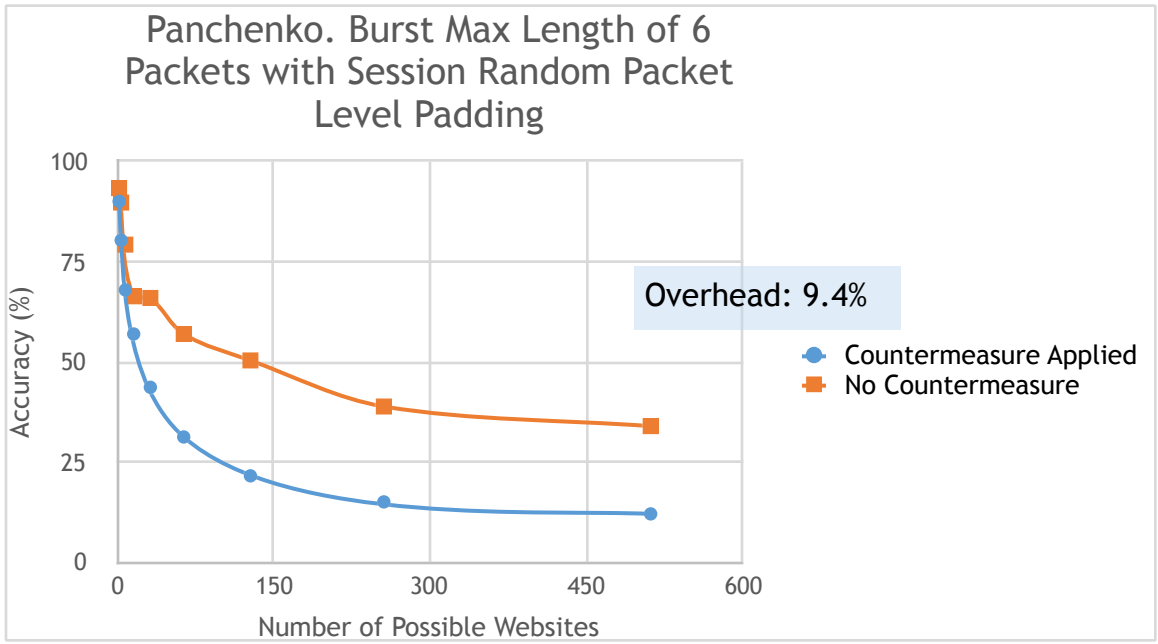
Burst Max Length of 6 Packets with Session Random Packet Level Padding

No Countermeasure:

Parameters:
P = 88

Countermeasure Applied:



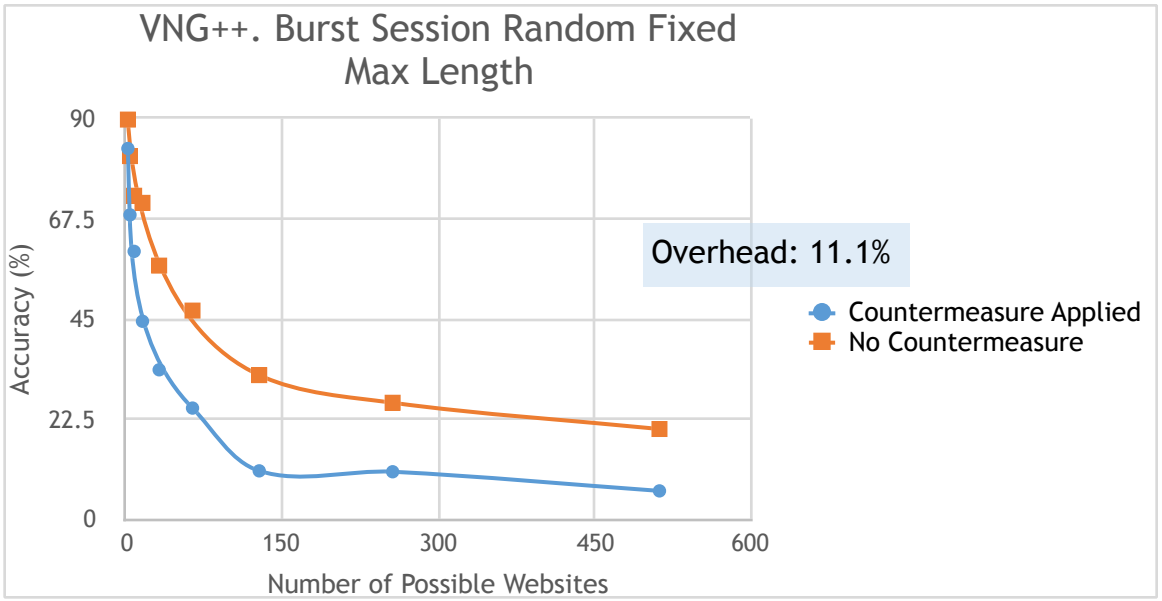


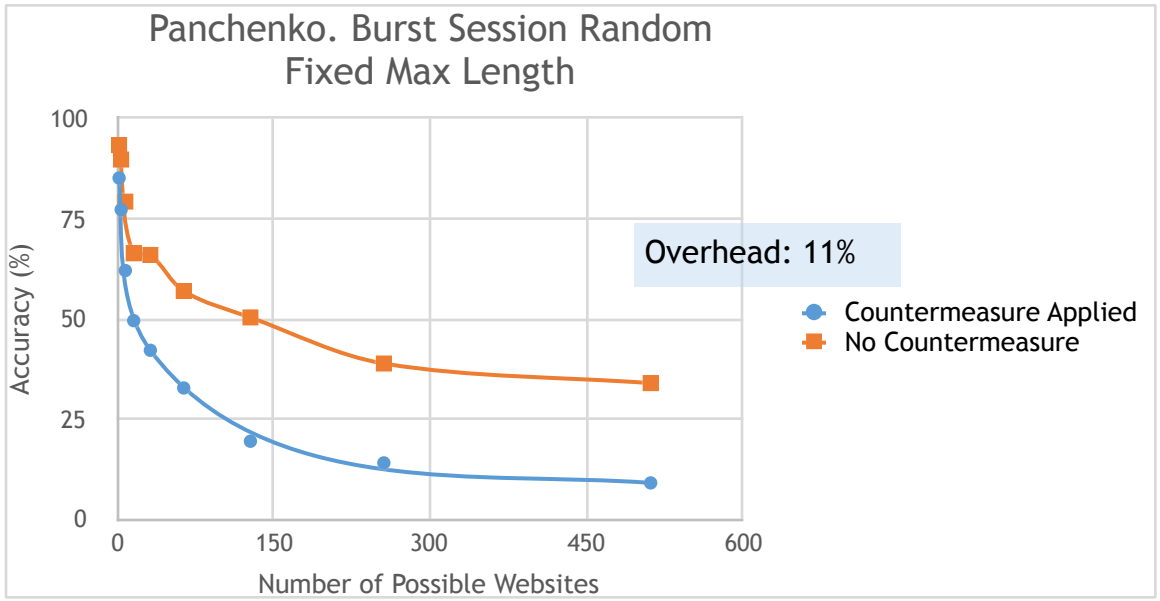
Burst Session Random Fixed Max Length

No Countermeasure:

Parameters:
P = 88
M = 7

Countermeasure Applied:





Burst Fixed Max Length

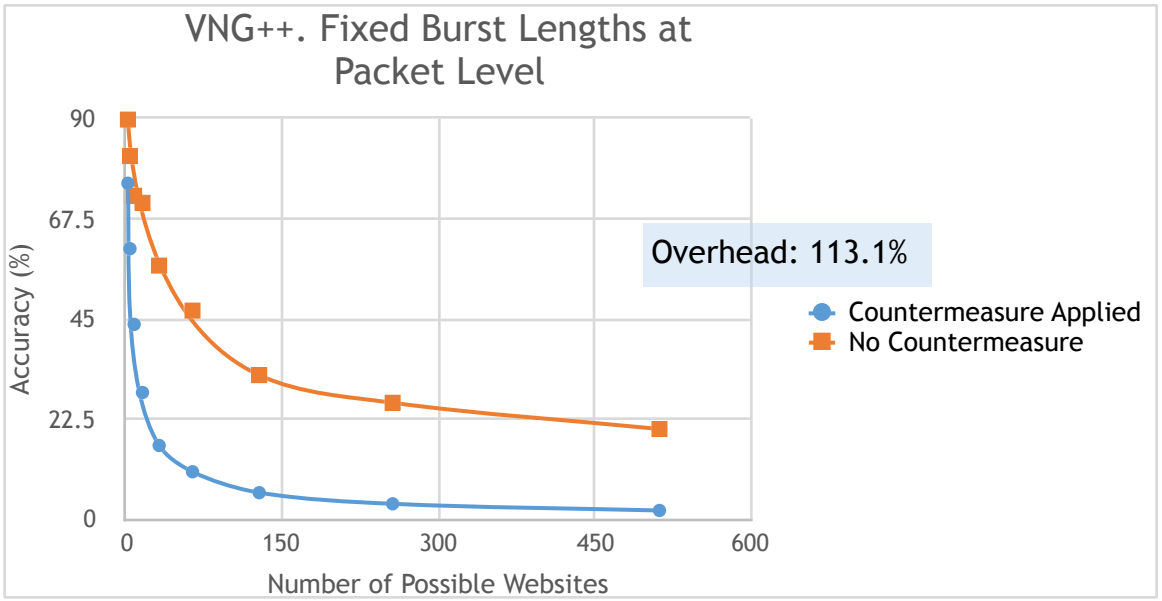
Advantages	Disadvantages
• Low Overhead • • Effectively mask the bursts when randomness is added • • Decreases accuracy of VGN++ from 20.3% • •	• Ineffective without packet level padding as well • Does not increase effectiveness of Session Random 255 with Panchenko classifier

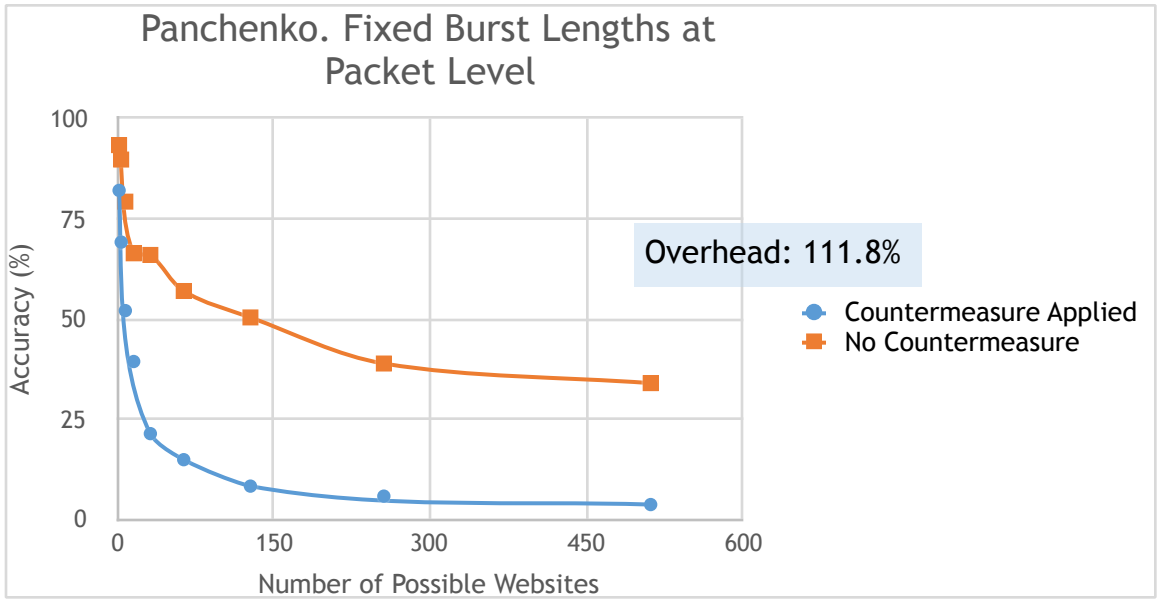
Fixed Burst Lengths at Packet Level

No Countermeasure:

Parameters:
P = 128
B = 5

Countermeasure Applied:



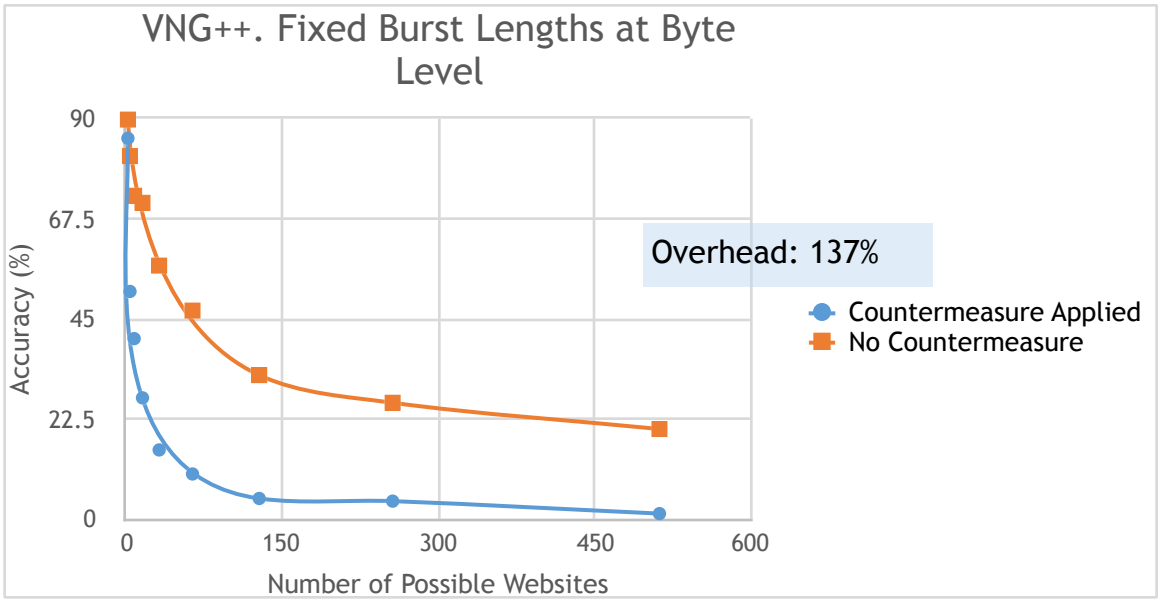


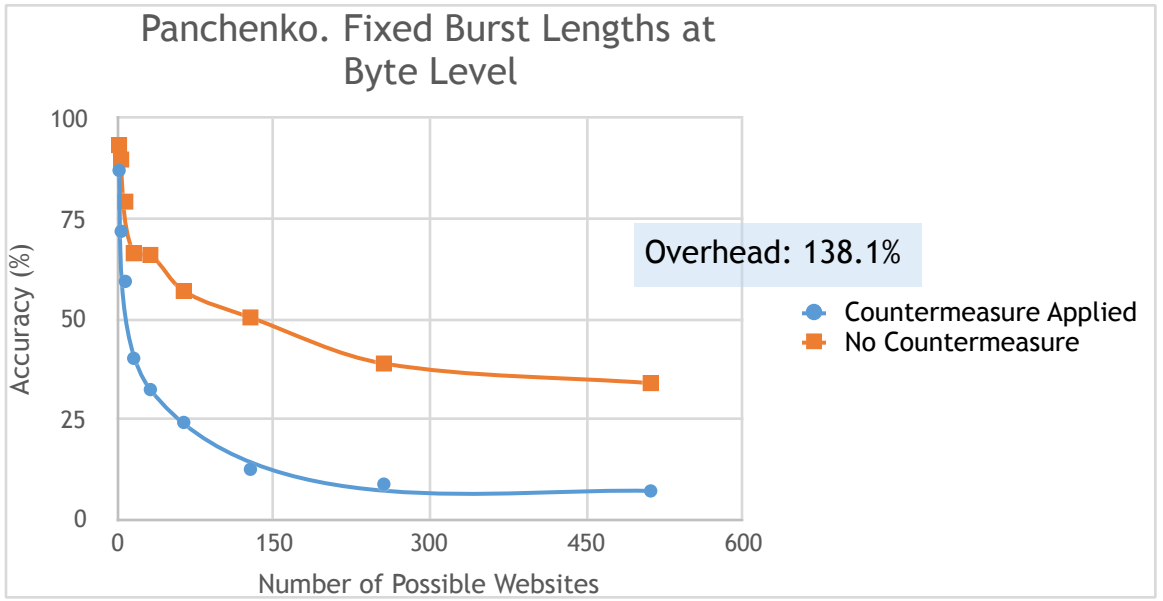
Fixed Burst Lengths at Byte Level

No Countermeasure:

Parameters:
P = 144
G = 9118

Countermeasure Applied:





Burst Fixed Length

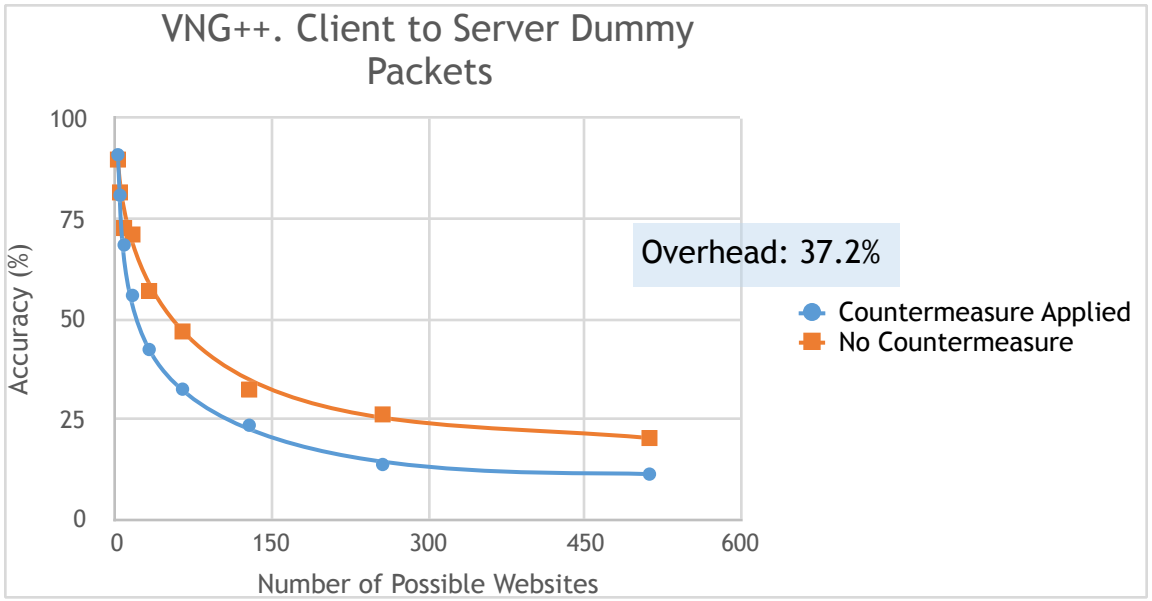
Advantages	Disadvantages
• Both implementations achieve low accuracy against VNG++ • • Packet level implementation achieves low accuracy against Panchenko • •	• High Byte Overhead- takes more than twice as long for a web page to download • • Potential for high latency overhead

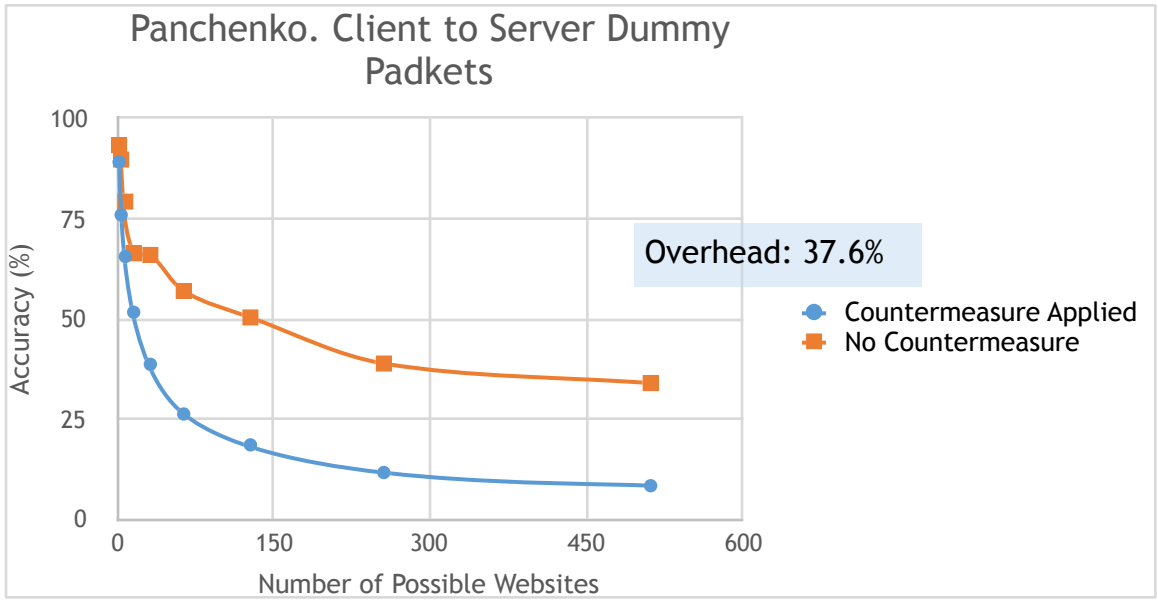
Client to Server Dummy Packets

No Countermeasure:

Parameters:
P = 80
B = 4
C = 3

Countermeasure Applied:

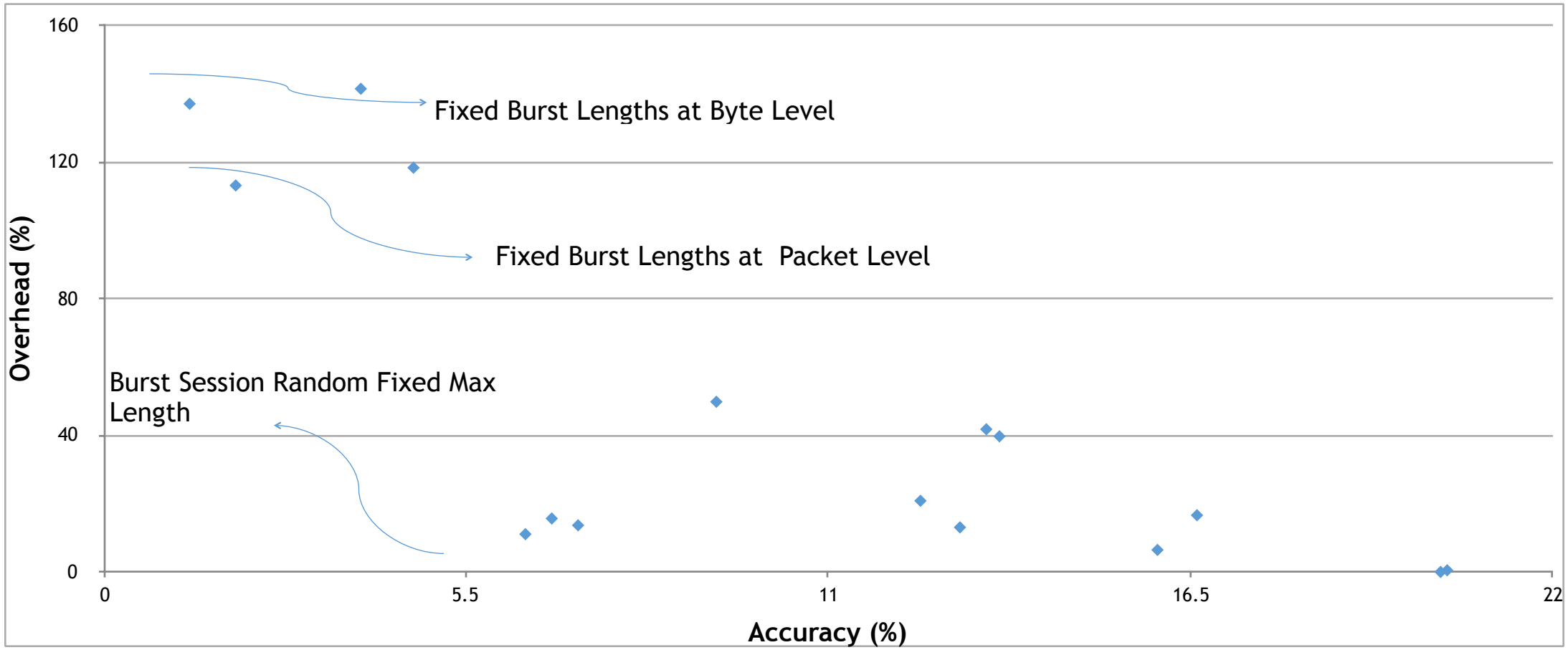




Client to Server Dummy Packets

Advantages	Disadvantages
• Low byte level overhead • • Possibly take advantage of asymmetric bandwidth • •	• Cannot achieve accuracy results as strong as Server to Client Dummy Packets • • Potential for high latency overhead •

Summary: Accuracy vs. Overhead Tradeoff



Conclusion

- Introduced **Strategic Dummy Packet Placement- Burst Level Padding**

- **Decrease accuracy** of most effective countermeasure, Session Random 255, from **16.6% to 6.4%** with Burst Fixed Max Length
  - Slight overhead increase from 6.6% to 11.1%

- **Decrease accuracy** of most effective countermeasure, Session Random 255, from **16.6% to 1.3%** with Fixed Burst Lengths at Byte Level
  - Large overhead increase from 6.6% to 137%

Opportunities for Future Research

- **Common Characteristics of a Trace:** Application of countermeasure on select parts of trace
  - Taking advantage of “anatomy of a trace”
- **Taking advantage of User Habits:** Eliminating “open-world concerns” and increasing false positives

Acknowledgements

- Our advisor, Professor Prateek Mittal
- Kevin Dyer, for access to his code base and support in experiment recreations
- Xiang Cai, for access to his code base and support in experiment recreations

Citations

[1] Andriy Panchenko, Lukas Niessen, Andreas Zinnen, and Thomas Engel. Website Fingerprinting in Onion Routing based Anonymization Networks. In Proceedings of the Workshop on Privacy in the Electronic Society, pages 103–114, October 2011.

[2] Xiang Cai, Xincheng Zhang, Brijesh Joshi, and Rob Johnson. Touching from a distance: website fingerprinting attacks and defenses. In ACM CCS, 2012

[3] Kevin P. Dyer, Scott E. Coull, Thomas Ristenpart, and Thomas Shrimpton. Peek-a-boo, i still see you: Why e fficient traff c analysis countermeasures fail. In Proceedings of the 33rd Annual IEEE Symposium on Security and Privacy, 2012.