Analyzing Tor's Anonymity with Machine Learning

Sanjit Bhat, David Lu
May 21st, 2017
Mentor: Albert Kwon

Acknowledgements

Thank you to Albert Kwon for mentoring us

Thank you to Prof. Srini Devadas for PRIMES-CS

Thank you to Dr. Slava Gerovitch and the PRIMES program

Thank you to our parents for supporting us and for listening to us incessantly talk about our project for the past four months!

Motivation and Background

Anonymity Matters

- Whistleblowers
- Governmental suppression of political opinion
- Censorship circumvention

http://blog.transparency.org/2016/06/20/new-whistleblower-protection-law-in-france-not-yet-fit-for-purpose/

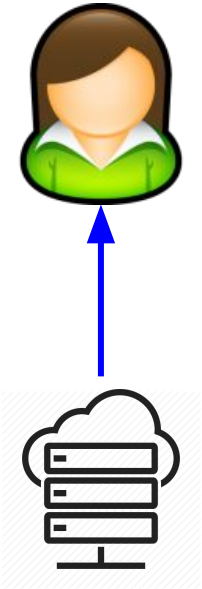
http://facecrooks.com/Internet-Safety-Privacy/To-be-anonymous-or-not-to-be-should-you-use-your-real-name-on-the-Internet.html/

http://www.dmnews.com/social-media/what-if-people-want-their-internet-anonymity-back/article/338654/

The Internet Provides No Guarantee of Anonymity

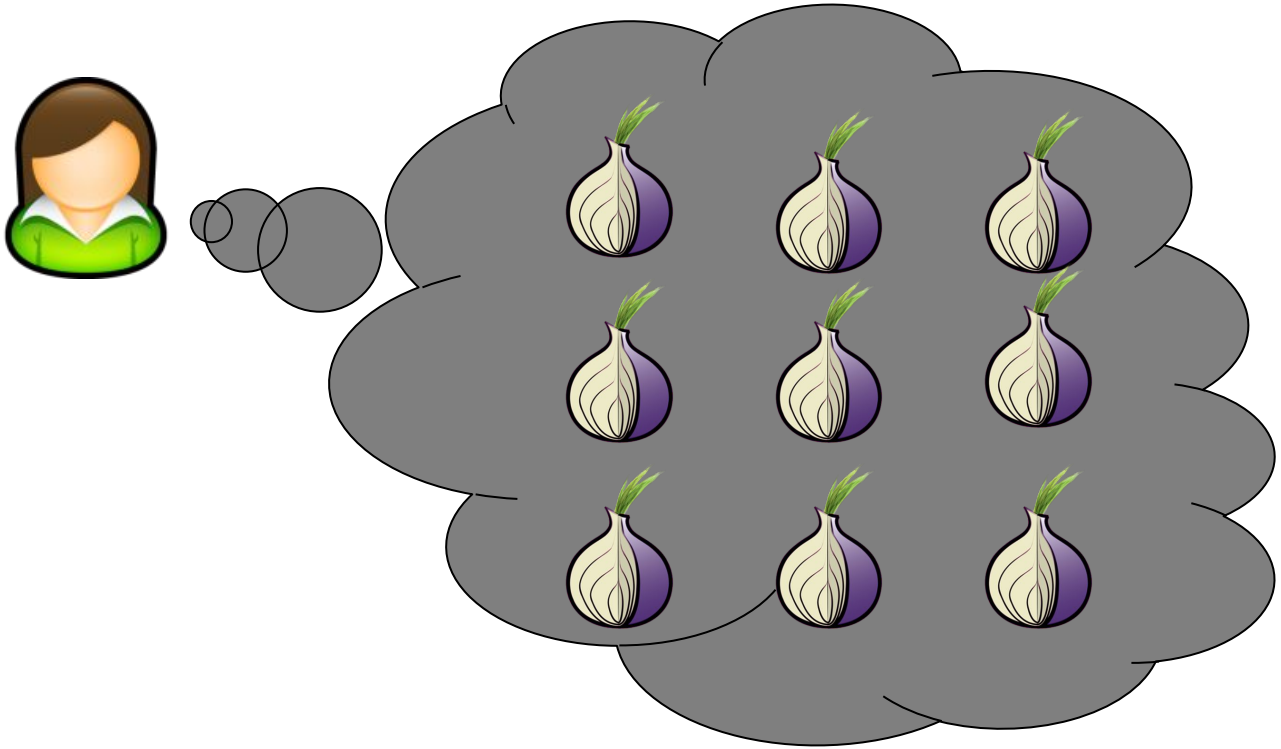
A Supposed Fix - Tor: The Onion Router

- Alice connects to the Tor network



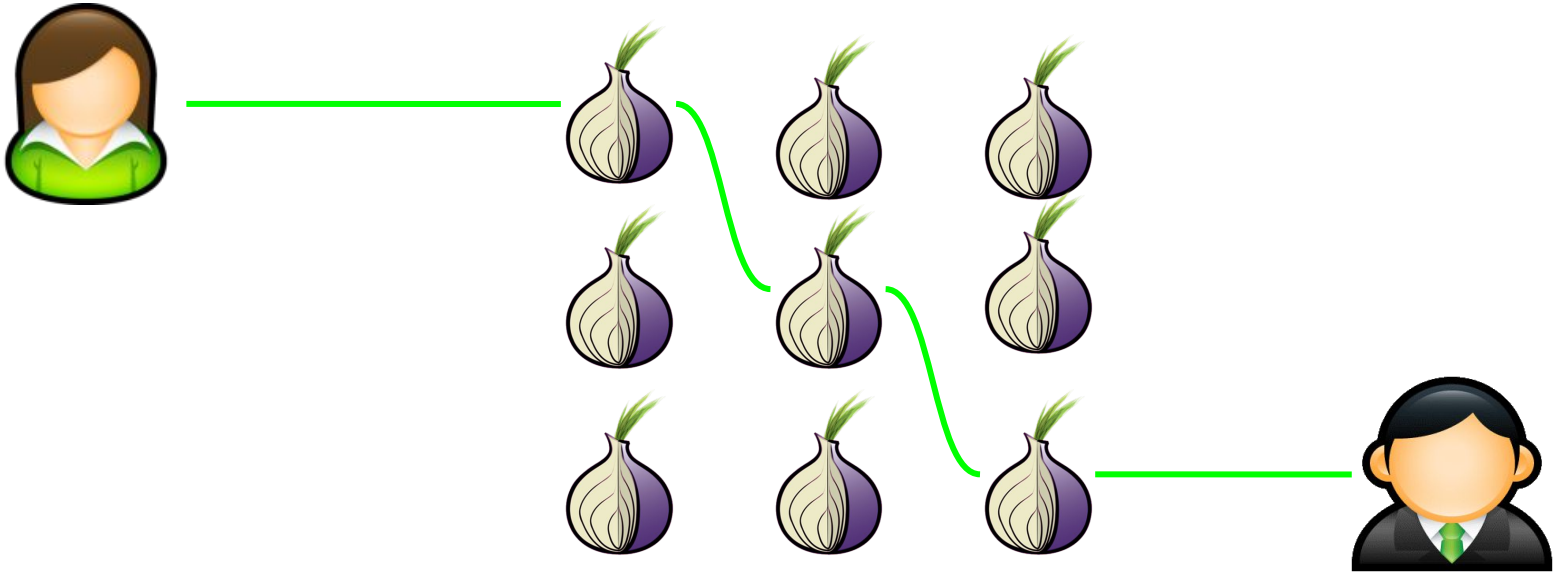
A Supposed Fix - Tor: The Onion Router

- Alice obtains a list of Tor nodes from the Tor network



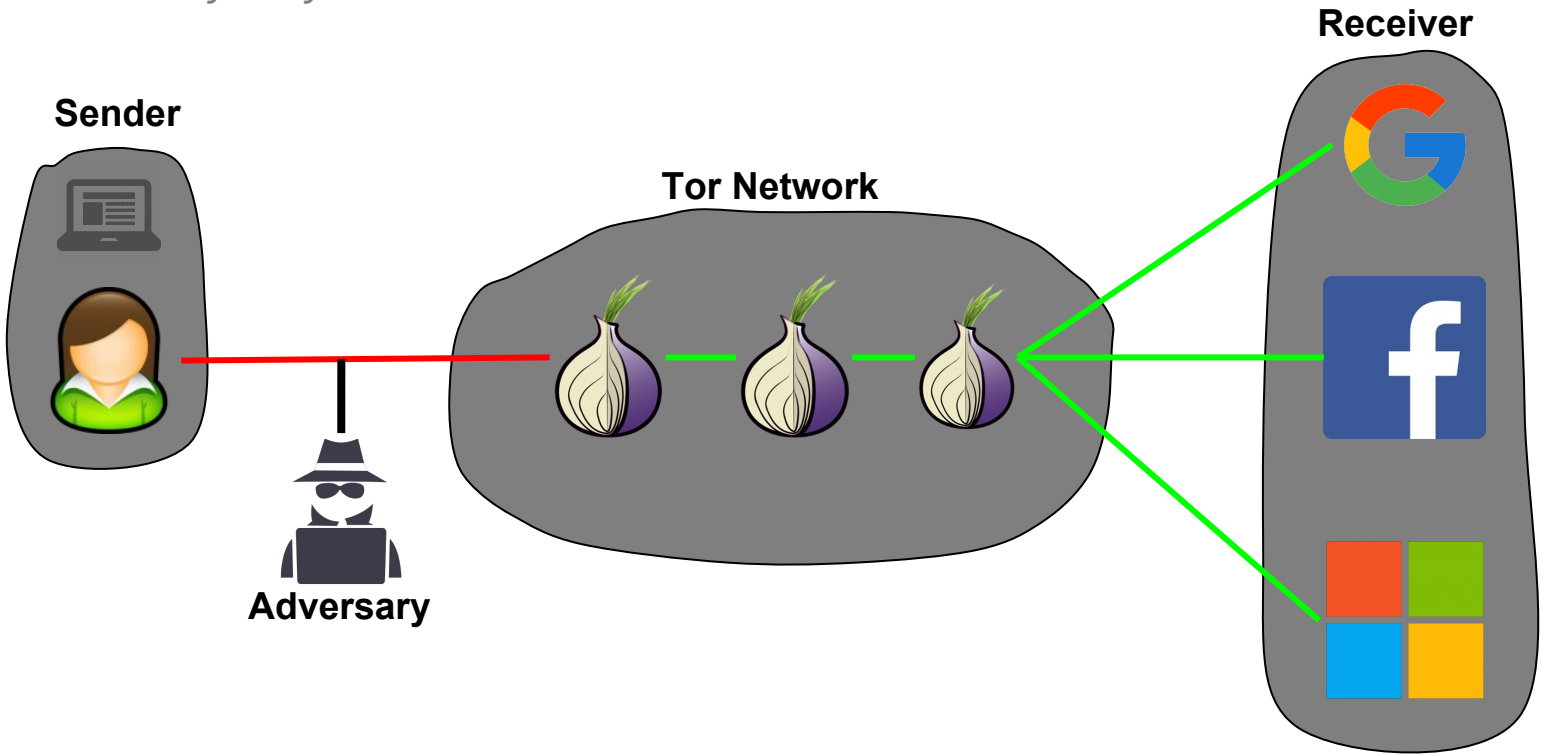
A Supposed Fix - Tor: The Onion Router

- Alice chooses 3 Tor nodes to make a connection to Bob
- No Tor nodes know the identities of both Bob and Alice



Tor Has Key Vulnerabilities Exploited by Website Fingerprinting Attacks

- Adversary only needs 1 link in the chain



Adversaries Learn the Following Features from Packet Sequences

Basic Features

- Total transmission time
- Total number of packets
- Number of incoming packets
- Number of outgoing packets

More Abstract Features

- Number of packets before outgoing packets
- Number of packets between outgoing packets
- Concentrations of outgoing packets
- Bursts
- Initial packet directions

T. Wang et al.: “Effective Attacks and Provable Defenses for Website Fingerprinting”

Our Research Focus - Applications of Website Fingerprinting to Unlocking Tor's Anonymity

Question 1: Are website fingerprinting attacks still viable?

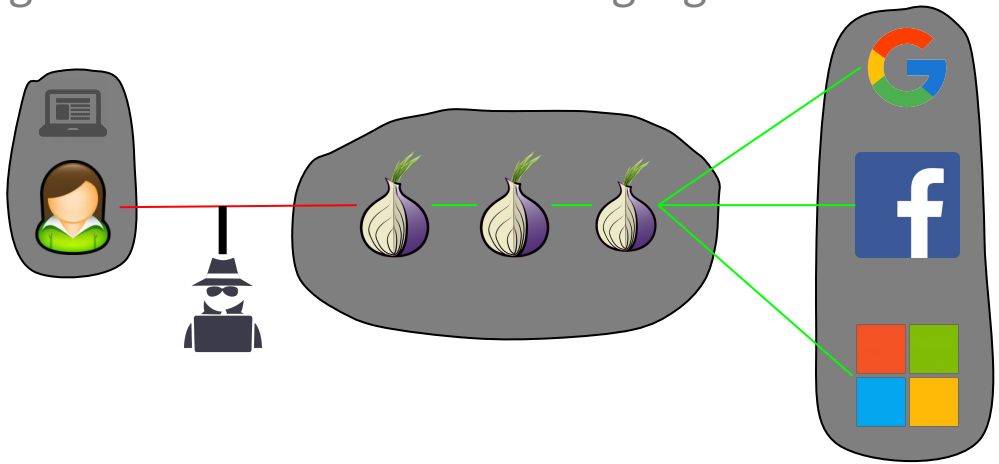
Question 2: Are certain features more important for the attacker?

Question 3: Can we classify websites based on their content type?

Experiments and Results

Methodology

- Our Dataset
  - Top 10,000 Sites from StuffGate
  - 50 trials per site using the Tor network
  - Mimic actions of actual user as closely as possible
- Evaluation
  - Multiple generic Weka machine learning algorithms



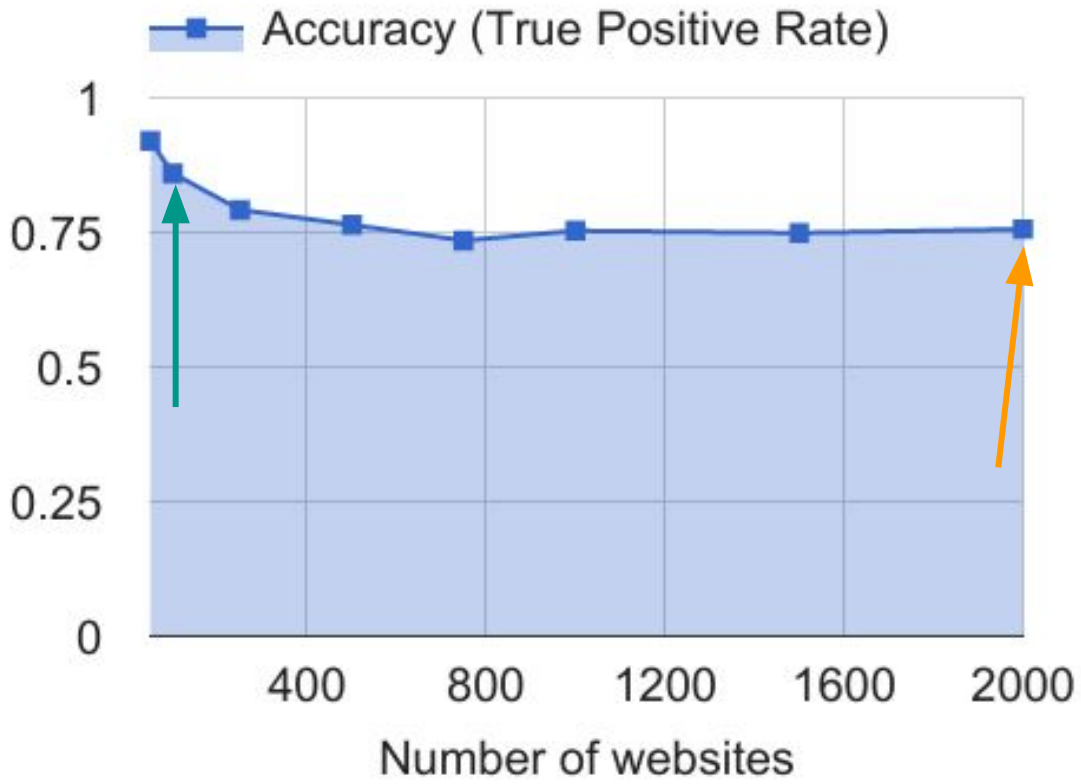
Question 1: Are website fingerprinting attacks still viable?

Closed World - A Necessary Benchmark

K-Nearest Neighbors Classifier

100 websites: 86%

2000 websites: 75.6%



Open World - A More Realistic Scenario

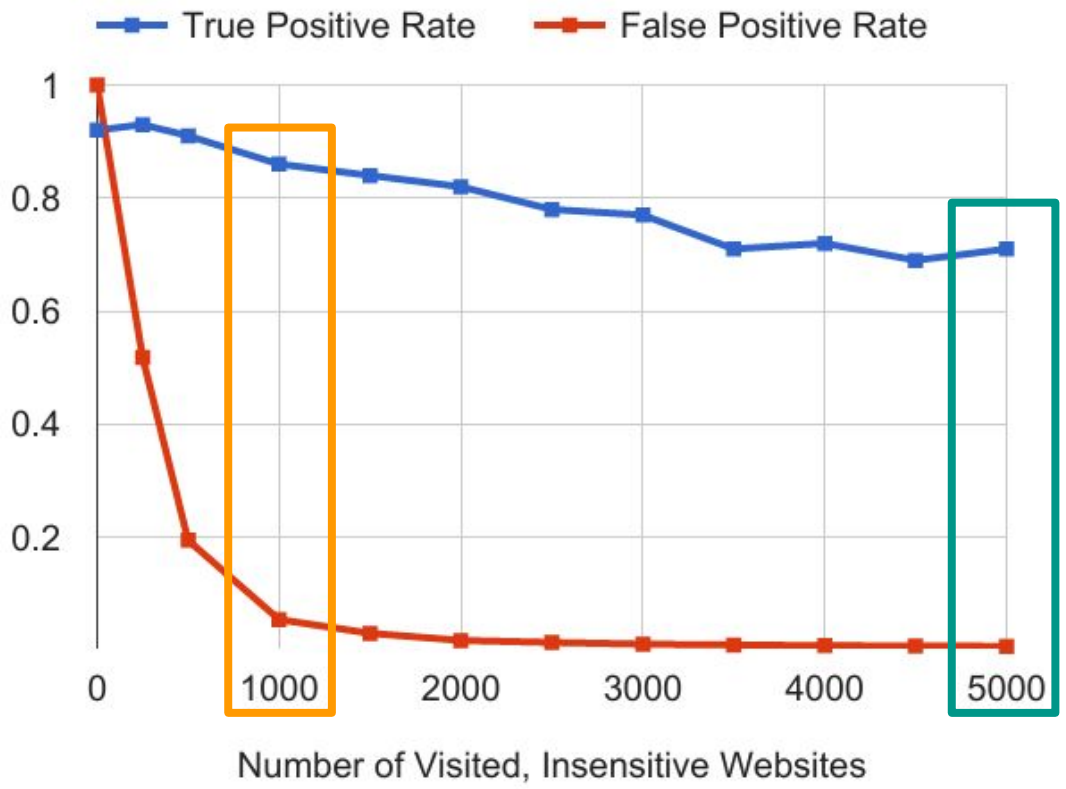
Random Forest Classifier

100 sensitive websites

5000 insensitive websites

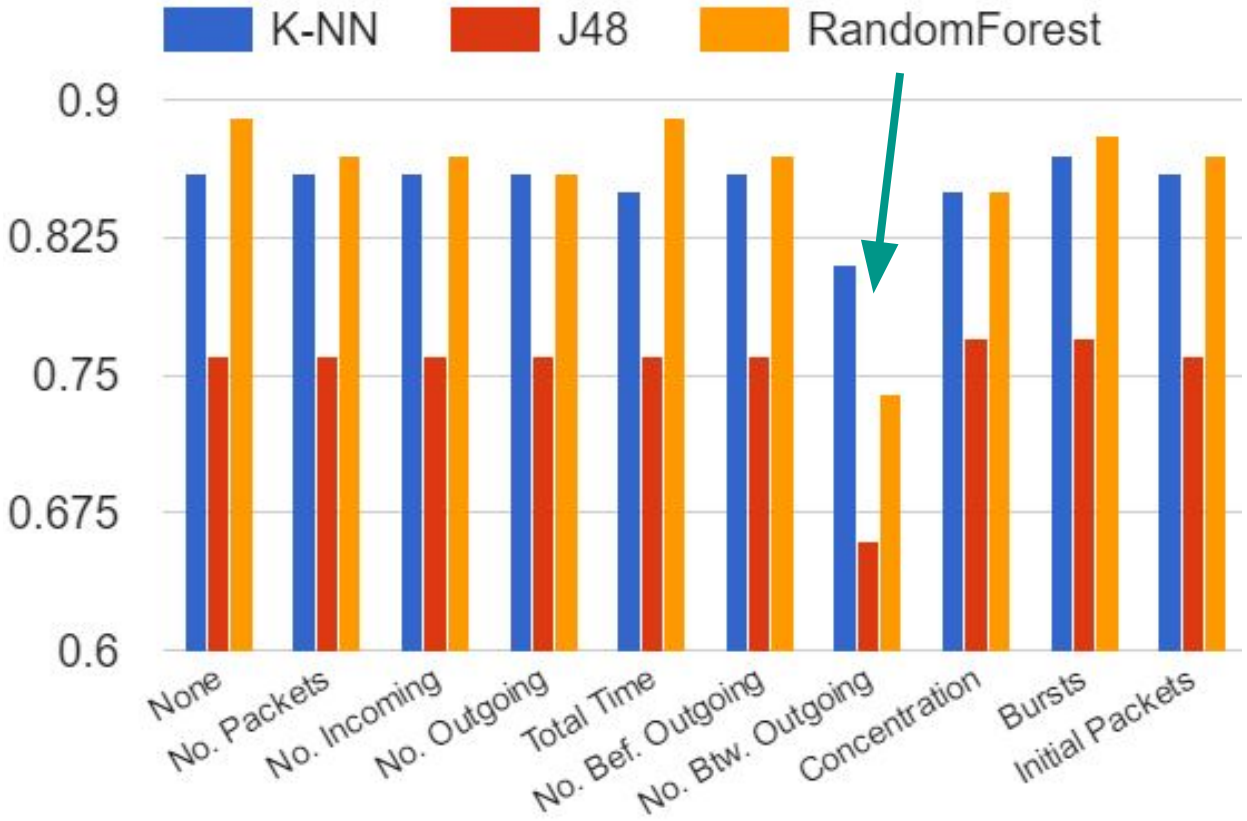
True Positive Rate: 86%
False Positive Rate: 5.5%

True Positive Rate: 71%
False Positive Rate: 0.76%



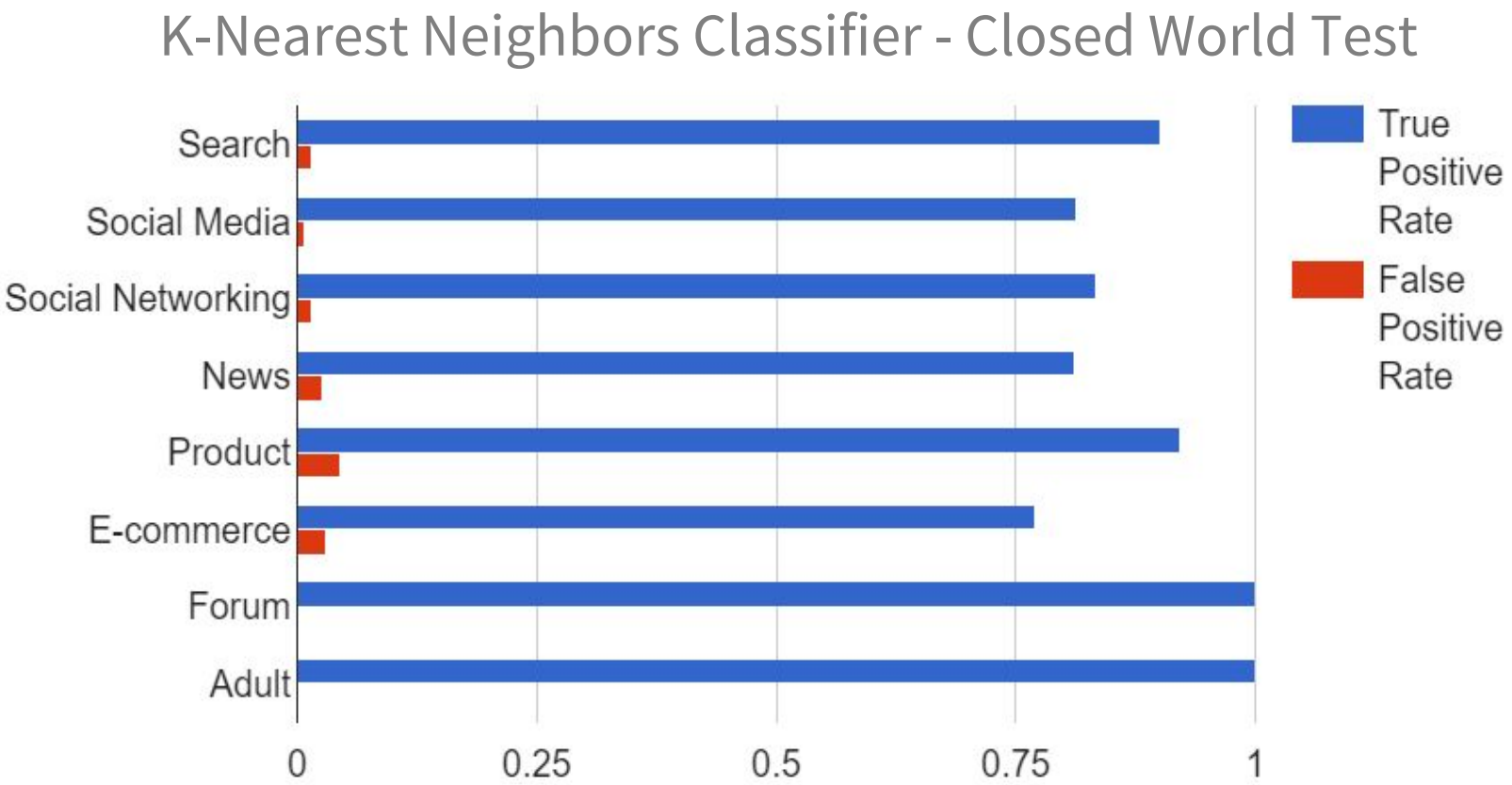
Question 2: Are certain features more important for the attacker?

Packet Ordering - The Most Impactful Feature



Question 3: Can we classify websites based on their content type?

Websites Can Be Classified into Categories



Conclusion

- Website fingerprinting attacks are still viable and can be used to circumvent Tor's anonymity
- Among the features available to an adversary observing the Tor network, the number of incoming packets between successive outgoing packets is the most important feature
- Adversaries can successfully classify, by content type, the websites accessed by users on the Tor network

Future Work

- Attacks
  - More powerful/tailored machine learning algorithms
    - Scikit-learn
  - Deep-learning
    - Capture abstract features not humanly visible
- Defenses
  - Hide packet ordering
  - Unsupervised learning
    - Cluster sites together to find categories not humanly identifiable