Poster: Encrypted Network Traffic Analysis

Madeline Moran, Joshua Honig, Nathan Ferrell, Shreena Soni
Sophia Homan, Eric Chan-Tin, Mohammed Abuhamad
Loyola University Chicago
Contact: mmoran11@luc.edu and jhonig@luc.edu

***Abstract*—Maintaining ones privacy online is often thought of as using strong passwords and using encrypted network communications. However, website fingerprinting has in the past been proven to expose even encrypted and anonymous communications. Using packet size, direction, and the number of network packets, a Website Fingerprinting adversary can accurately predict websites a user visits online. In this research, we go beyond the typical website fingerprinting attack and show an adversary is able to identify specific articles a user visits, specific Google Searches, and specific actions they take in a virtual reality system. We analyzed encrypted network traffic and used a RandomForest classification machine learning algorithm to predict the specific action a user is undertaking. For the virtual reality system, our system obtained a 91% accuracy in identifying the action taken and for the webpage fingerprinting, our system obtained a 70% accuracy.**

***Index Terms*—Network Traffic Analysis, Website Fingerprinting, Privacy, Traffic Interception**

I. Introduction

It is impossible to function today without using the internet. We work, interact with friends, consume content, and manage our lives online, making our internet traffic incredibly valuable. Eavesdroppers have long been interested in intercepting internet traffic to either pull out sensitive information or analyze user behavior. To combat this, SSL/TLS encryption is now used by default.

Thus, determining user behavior from encrypted network traffic is important for many parties. For example, law enforcement may be interested to know if a user is visiting websites known for criminal activity, or corporate network administrators may desire to know if users are visiting websites which violate their policy. Such actors can make these determinations using a Website Fingerprinting attack (WF), removing the need to decrypt the traffic. WF is the practice of using the metadata of encrypted network traffic to build a database of known network traffic patterns for specific websites, called fingerprints. Then, the attacker compares unknown network traffic to those fingerprints to predict which website a user is visiting. This is generally achieved with machine learning.

II. Threat Model

Figure 1 shows our threat model with an adversary who is able to read all network communications to and from a specific user via an internet connection. They are unable to read any encrypted information or the destination IP address but they are able to collect and read metadata such as packet sizes and packet direction across the network. The goal of the adversary is to identify the website that the user is visiting.

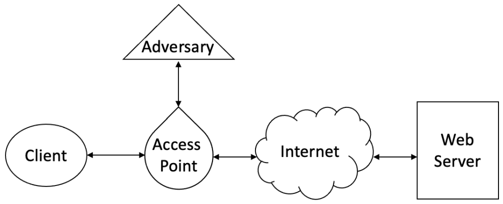


Fig. 1. Threat Model.

III. Experimental Design

We conducted three different kinds of website fingerprinting attacks: 1) Webpage Fingerprinting, 2) Google Autocomplete and Search fingerprinting, and 3) Virtual Reality fingerprinting. Each experiment was designed to collect network traffic for multiple instances and compare them using a machine learning algorithm. The goal of each experiment is to train the model to correctly identify the users' "action" with a high degree of accuracy based on the network traffic, with "action" representing the webpage, keyword search, or VR task.

A. Experiments

To begin each experiment, we created an automated process to send thousands of page requests and monitor network traffic to generate a sample size for each action. For the VR sample, we manually created and acted out a list of tasks, such as playing a game, login/logout of Netflix, perform search queries on different platforms such as a map, Google, and Amazon.

Our first experiment focused specifically on identifying articles (webpage fingerprinting) from the same source rather than a multitude of unrelated webpages seen in traditional WF. The goal was to see if we could identify fingerprints, even within the same website. To this end, we analyzed two popular websites, Wikipedia and New York Times (NY Times). We hypothesized that Wikipedia would have less significant deviations in their network traffic due to the uniformity of their articles and underlying website structure. In contrast, we hypothesized NY Times articles would be more complex and dynamic given their varied stylistic choices and the additional traffic generated by things like website trackers. We collected $1,000$ webpages and 20 samples of each webpage.

Our next experiment aims to fingerprint network traffic as a result of Google Autocomplete activity and their corresponding Search results. These two activities were collected together but analyzed separately. The goal of this experiment was to test if the network activity before a website was even selected could be identified to better understand what a user might be looking for. We collected $1,144$ searches and 25 samples of each query.

Our final experiment analyzed the network traffic resulting from Virtual Reality (VR) user activity. Our goal for this experiment was to see if even one of the newer forms of popular technology was susceptible to network traffic fingerprinting. We collected network data for 20 actions and 5 samples of each action.

B. Data Processing

We decided to analyze our network traffic data further by looking for identifying features packet traffic. This process was meant to summarize the data as well as display only a certain number of packets in order to refine what the machine learning algorithm would analyze. After our chosen algorithm, RandomForest, was selected, we performed a 5-fold cross-validation and grid-search to find the optimal hyperparameters.

IV. Results

For our webpage fingerprinting experiment, we were able to identify the individual Wikipedia article 69.85% of the time and the NY Times article 48.00% of the time, as seen in Figure I. This confirms our hypothesis that the Wikipedia pages would be uniform without the need to load in advertisements and a lack of complex page formatting. In contrast, the NY Times articles would be more difficult to identify than the Wikipedia pages due to the excess of information and formatting that isn't necessarily relevant to the article but is necessary to render each page.

TABLE I
Results.

Experiment	***Accuracy***	***# of Classes***
Wikipedia	69.85%	1000
New York Times	48.0%	1000
Google Autocomplete	21.63%	1143
Google Search	15.41%	1143
Virtual Reality	90.91%	14

For our Google Autocomplete and Search results we only obtained an accuracy of 21.63% for Autocomplete results and 15.41% for Search results. Despite this, our accuracy is still significant because a random model could have only received an accuracy of $1/1,143$ samples $= 0.088\%$. Our model obtained an accuracy over 172 times the base accuracy. This shows that even though Google search results are similar when viewed as a webpage, information can still be extracted from the network traffic to predict the exact Google search query performed by a user.

Our VR experiment received an accuracy of 90.91% over 14 classes. While the high accuracy is in part due to the number of classes we had as a result of the manual data collection process, it is still significant that we found that network traffic fingerprinting could be performed on a VR system.

V. Related Work

Numerous features, including packet tinting [1], cumulative sizes [2], and n-grams [3], and machine learning algorithms, including SVM (Support Vector Machine), k-NN (k-Nearest-Neighbor), and deep learning [4], have been used to predict websites. Additionally, some experiments have used non-website applications such as social media apps [5], voice recognition [6], web searches [7], and DNS over HTTPS [8], which corroborates our hypothesis that website fingerprinting can be conducted over multiple applications.

VI. Future Work

In terms of future work, it is likely capturing more instance of each action would provide more insight into each fingerprint and raise the accuracy. For the VR experiment, future work could attempt to predict the exact keyword search, rather than the action of searching. A more thorough grid-search could also be performed to obtain the best hyperparameters for the best accuracy.

VII. Acknowledgements

We thank Taryn Chovan for helping with the code implementation and Taylor Mansheim for providing feedback on our research. Project sponsored by the National Security Agency under Grant Number H98230-21-1-0325. The United States Government is authorized to reproduce and distribute reprints not withstanding any copyright notation herein. This manuscript is submitted for publication with the understanding that the United States Government is authorized to reproduce and distribute reprints. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Security Agency.

References

[1] J. Juen, A. Johnson, A. Das, N. Borisov, and M. Caesar, "Defending tor from network adversaries: A case study of network path prediction," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 2, pp. 171–187, 2015.

[2] A. Panchenko, F. Lanze, A. Zinnen, M. Henze, J. Pennekamp, K. Wehrle, and T. Engel, "Website fingerprinting at internet scale," in *Proceedings of the 23rd Internet Society (ISOC) Network and Distributed System Security Symposium (NDSS 2016)*, 2016.

[3] S. Siby, M. Juarez, C. Diaz, N. Vallina-Rodriguez, and C. Troncoso, "Encrypted DNS ==> Privacy? A Traffic Analysis Perspective," in *NDSS*, 2020.

[4] P. Sirinam, M. Imani, M. Juarez, and M. Wright, "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. ACM, 2018, pp. 1928–1943.

[5] M. Di Martino, P. Quax, and W. Lamotte, "Realistically fingerprinting social media webpages in https traffic," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ser. ARES '19, New York, NY, USA, 2019.

[6] J. Hyland, C. Schneggenburger, N. Lim, J. Ruud, N. Mathews, and M. Wright, "What a shame: Smart assistant voice command fingerprinting utilizing deep learning," in *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*, ser. WPES, 2021.

[7] S. E. Oh, S. Li, and N. Hopper, "Fingerprinting keywords in search queries over tor," *PoPETs*, vol. 2017, 2017.

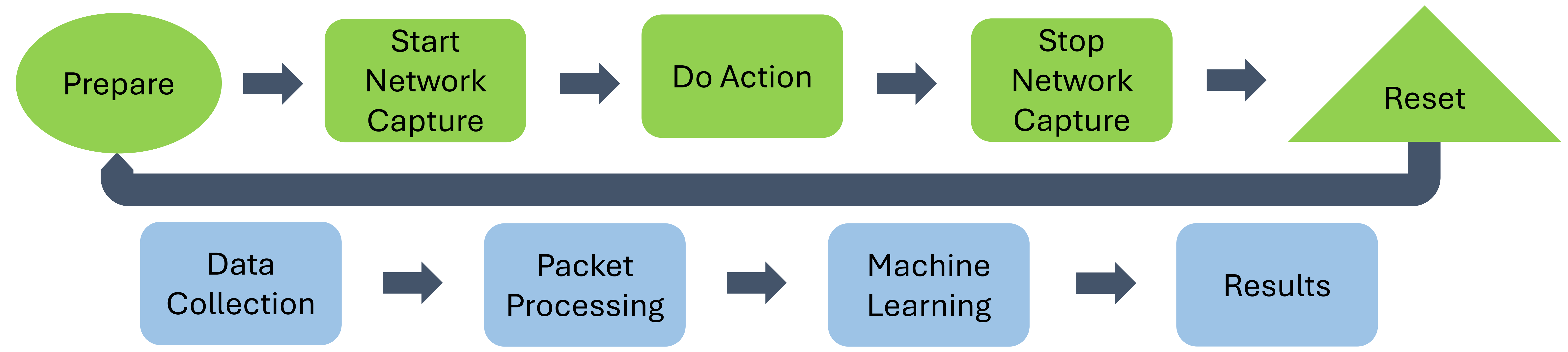
[8] D. Vekshin, K. Hynek, and T. Cejka, "Doh insight: Detecting dns over https by machine learning," in *Proceedings of the 15th International Conference on Availability, Reliability and Security*, 2020, pp. 1–8.

Encrypted Network Traffic Analysis

Madeline M. Moran, Joshua Honig, Nathan Ferrell, Shreena Soni,
Sophia Homan, Eric Chan-Tin, Mohammed Abuhamad

jhonig@luc.edu

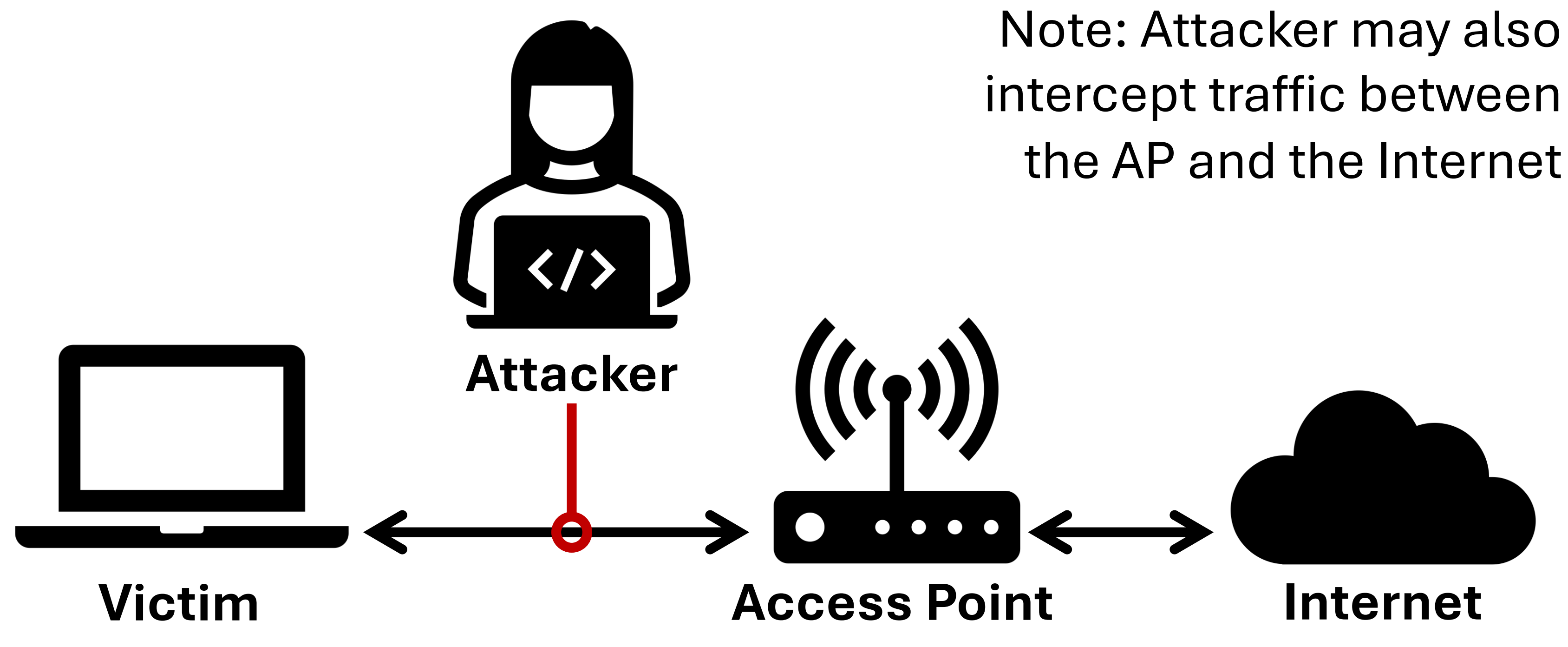
Procedure



Abstract

- User behavior is predictable from encrypted packet size and direction alone
- This study focuses on identifying specific web pages, Google searches, & actions on VR headset
- The RandomForest ML classification algorithm can be used to predict a user's online activity

Threat Model



Note: Attacker may also intercept traffic between the AP and the Internet

Gaining this level of visibility is trivial for an attacker

Hypotheses

Based only on encrypted network traffic metadata…

Wikipedia
…can we guess which Wikipedia article a user is visiting in the browser?

NY Times
…can we guess which NY Times article a user is visiting in the browser?

Google
…can we: guess what a user is typing in the search box; guess their query?

MetaQuest VR
…can we guess which actions a user takes on a MetaQuest VR headset?

Related Work

Social Media Apps [1] Voice Recognition [2] DNS over HTTPS [3]

References

[1] M. Di Martino, P. Quax, and W. Lamotte, "Realistically fingerprinting social media webpages in https traffic," in Proceedings of the 14th International Conference on Availability, Reliability and Security, ser. ARES '19, New York, NY, USA, 2019.

[2] J. Hyland, C. Schneggenburger, N. Lim, J. Ruud, N. Mathews, and M. Wright, "What a shame: Smart assistant voice command fingerprinting utilizing deep learning," in Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society, ser. WPES, 2021.

[3] S. E. Oh, S. Li, and N. Hopper, "Fingerprinting keywords in search queries over tor," PoPETs, vol. 2017.

Data Collection & Processing

Wikipedia, NY Times
- Collected 20 samples of 1,000 articles for each website

Google Search, Google Autocomplete
- Collected 25 samples of 1,143 queries
- Autocomplete: captured traffic for simulated query entry

MetaQuest VR
- Collected 5 samples of 14 user actions

Data Processing
- From each capture, extracted size and direction of each packet
- Then generated features: total, std. dev., average, max, min, etc. to understand the patterns of incoming and outgoing traffic

Data Classification (ML)

Dataset	Criterion	Maximum Depth	Maximum Features	Number of Estimators
Wikipedia	gini	80	None	400
NY Times	gini	None	None	500
Google Auto.	gini	110	sqrt	500
Google Search	gini	None	sqrt	500
VR Headset	entropy	None	sqrt	100

Highest Performing Hyperparameters

RandomForest Classifier & Cross Validation (CV)
- Better performance over AdaBoost, DecisionTree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN)
- CV allows training on all data, useful with low sample sizes
- Low barrier to entry (CPU & GPU libraries available)

Grid Search (GS)
- With 5-fold CV to determine best hyperparameters & increase accuracy; highest performing hyperparameters in table above

Conclusions & Limitations

Dataset	Accuracy	Samples
Wikipedia	69.85%	1000
NY Times	48.00%	1000
Google Autocomplete	21.63%	1143
Google Search	15.41%	1143
VR Headset Actions	90.91%	14

random guess: 0.1% (Wiki, NYT); 0.0875% (Google); 7.143% (VR)

Limitations
- Collection timeframe was long; content may change
  - Especially problematic for Google Search dataset
- Low sample size, especially for VR

Conclusions
- VR, Wikipedia, NY Times fingerprintable
- Google most fingerprintable by client request traffic

Acknowledgement

We thank Taryn Chovan for helping with the code implementation and Taylor Mansheim for providing feedback on our research. Project sponsored by the National Security Agency under Grant Number H98230-21-1-0325. The United States Government is authorized to reproduce and distribute reprints not withstanding any copyright notation herein. This manuscript is submitted for publication with the understanding that the United States Government is authorized to reproduce and distribute reprints. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Security Agency.