Optimization for Attack Surface Exploration

The Case of TLS

Vasilios Mavroudis
University College London

Jamie Hayes
University College London

Attack Surface Exploration

Given a system or protocol:

1. Study its operation, security properties and trust assumptions.
2. Define the adversaries and their goals.
3. Design, realise and optimize the attack(s).

Attack Surface Exploration

Given a system or protocol:

1. Study its operation, security properties and trust assumptions.
2. Define the adversaries and their goals.
3. Design and realise the attack(s).

Step 3 can be a complex and time-consuming iterative process.

Optimizing an Attack

Goal

Advance the state-of-the-art so as to better evaluate the security of the system/protocol and the available defences.

Optimizing an Attack

Manual iterative process

- How efficient is it? False positives/False negatives

- Are its assumptions realistic? Can we relax any?

- Can we do better?

Optimizing an Attack

Manual iterative process

- How efficient is it? False positives/False negatives

- Are its assumptions realistic? Can we relax any?

- Can we do better?

Best-effort process that requires human involvement.

Exploration with ML/AI

- Further advance the state-of-the-art.
- Minimize human involvement.

Exploration with ML/AI

- Further advance the state-of-the-art.
- Minimize human involvement.

Requirements

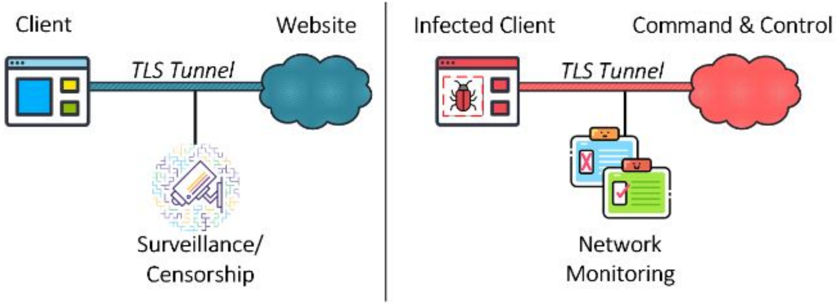
- Must be easier to collect data and train a model than to optimize the attack manually.

- The goal of the adversary can be expressed as a differentiable loss function.

Use Case: Traffic Fingerprinting

Webpage Fingerprinting (TLS)
Infer the *webpage* visited from the patterns of traffic between the client and the server.

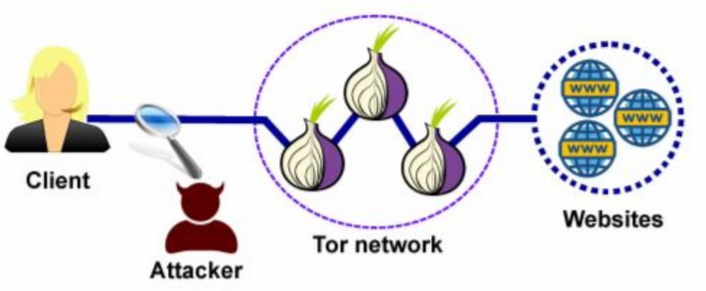
TLS does not protect the IP of the server visited.



Use Case: Traffic Fingerprinting

Website Fingerprinting (Tor)

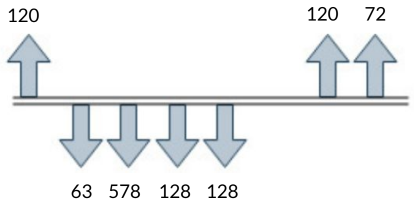
Infer the *website* visited from the traffic between the client and the Tor entry node.



Sirinam, Payap, et al. "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning." ACM Conference on Computer and Communications Security. 2018.

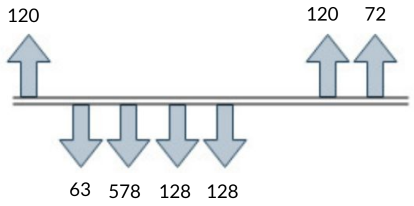
The Fingerprinting Task

- Traffic to-be-fingerprinted is encoded as sequences of bytes.



The Fingerprinting Task

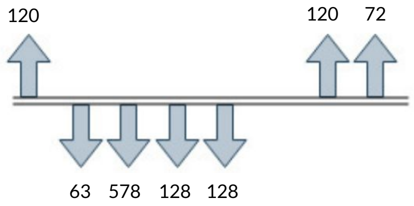
- Traffic to-be-fingerprinted is encoded as sequences of bytes.



1. The adversary compiles a dataset of labelled sequences.

The Fingerprinting Task

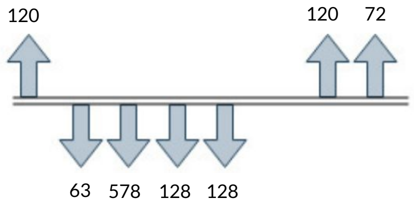
- Traffic to-be-fingerprinted is encoded as sequences of bytes.



1. The adversary compiles a dataset of labelled sequences. Easy to automate!

The Fingerprinting Task

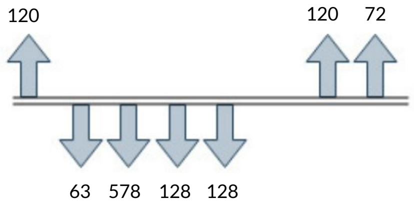
- Traffic to-be-fingerprinted is encoded as sequences of bytes.



1. The adversary compiles a dataset of labelled sequences. Easy to automate!
2. Prepares a classification system.

The Fingerprinting Task

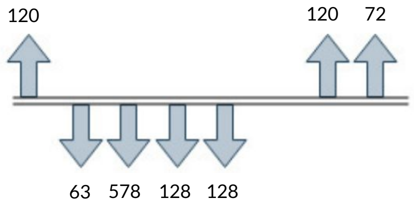
- Traffic to-be-fingerprinted is encoded as sequences of bytes.



1. The adversary compiles a dataset of labelled sequences. Easy to automate!
2. Prepares a classification system.
3. Classifies unknown sequences captured from a victim's traffic.

The Fingerprinting Task

- Traffic to-be-fingerprinted is encoded as sequences of bytes.



1. The adversary compiles a dataset of labelled sequences. Easy to automate!
2. Prepares a classification system.
3. Classifies unknown sequences captured from a victim's traffic.

Can modern ML help with steps 2 and 3?

...2016 | 2017-2018 | 2019-2020

...2016

2017-2018

2019-2020

Not an exhaustive review of the literature!

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. “Website fingerprinting in onion routing based anonymization networks.” WPES 2011.

Dyer, Kevin P., et al. “Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail.” IEEE S&P 2012.

Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014.

Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. "Website fingerprinting in onion routing based anonymization networks." WPES 2011.

Dyer, Kevin P., et al. "Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail." IEEE S&P 2012.

Wang, Tao, et al. "Effective attacks and provable defenses for website fingerprinting." USENIX 2014.

Cai, Xiang, et al. "A systematic approach to developing and evaluating website fingerprinting defenses." CCS 2014.

- Very good performance

Rimmer, Vera, et al. "Automated website fingerprinting through deep learning." Arxiv 2017/NDSS 2018

Sirinam, Payap, et al. "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning." CCS 2018.

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. “Website fingerprinting in onion routing based anonymization networks.” WPES 2011.

Dyer, Kevin P., et al. “Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail.” IEEE S&P 2012.

Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014.

Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.

- Very good performance

Rimmer, Vera, et al. “Automated website fingerprinting through deep learning.” Arxiv 2017/NDSS 2018

Sirinam, Payap, et al. “Deep fingerprinting: Undermining website fingerprinting defenses with deep learning.” CCS 2018.

“The success of such attacks heavily depends on the particular set of traffic features that are used to construct the fingerprint. Typically, these ***features are manually engineered****...”*

...2016	2017-2018	2019-2020
Incremental Steps: **- Feature Engineering** **- Performance**	**- Very good performance**	
Panchenko, Andriy, et al. "Website fingerprinting in onion routing based anonymization networks." WPES 2011. Dyer, Kevin P., et al. "Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail." IEEE S&P 2012. Wang, Tao, et al. "Effective attacks and provable defenses for website fingerprinting." USENIX 2014. Cai, Xiang, et al. "A systematic approach to developing and evaluating website fingerprinting defenses." CCS 2014.	**Rimmer, Vera, et al. "Automated website fingerprinting through deep learning." Arxiv 2017/NDSS 2018** Sirinam, Payap, et al. "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning." CCS 2018.	*"The success of such attacks heavily depends on the particular set of traffic features that are used to construct the fingerprint. Typically, these **features are manually engineered**..."* *"... we show that an adversary can **automate the feature engineering** process, and thus automatically deanonymize Tor traffic by applying our novel method based on deep learning."*

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. "Website fingerprinting in onion routing based anonymization networks." WPES 2011.

Dyer, Kevin P., et al. "Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail." IEEE S&P 2012.

Wang, Tao, et al. "Effective attacks and provable defenses for website fingerprinting." USENIX 2014.

Cai, Xiang, et al. "A systematic approach to developing and evaluating website fingerprinting defenses." CCS 2014.

- Very good performance
- Vulnerability approximators

Cherubin, Giovanni. "Bayes, not naïve: Security bounds on website fingerprinting defenses." PoPETs 2017

Rimmer, Vera, et al. "Automated website fingerprinting through deep learning." Arxiv 2017/NDSS 2018

Sirinam, Payap, et al. "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning." CCS 2018.

*"**Derive security bounds** for any WF defense, where the bounds depend on a chosen feature set."*

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. “Website fingerprinting in onion routing based anonymization networks.” WPES 2011.

Dyer, Kevin P., et al. “Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail.” IEEE S&P 2012.

Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014.

Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.

- Very good performance
- Vulnerability approximators

Cherubin, Giovanni. “Bayes, not naïve: Security bounds on website fingerprinting defenses.” PoPETs 2017

Rimmer, Vera, et al. “Automated website fingerprinting through deep learning.” Arxiv 2017/NDSS 2018

Sirinam, Payap, et al. “Deep fingerprinting: Undermining website fingerprinting defenses with deep learning.” CCS 2018.

“Derive security bounds for any WF defense, where the bounds depend on a chosen feature set.”

“...Such error can be estimated in practice, and is a ***lower bound for a WF adversary****, for any classification algorithm he may use.”*

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. "Website fingerprinting in onion routing based anonymization networks." WPES 2011.

Dyer, Kevin P., et al. "Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail." IEEE S&P 2012.

Wang, Tao, et al. "Effective attacks and provable defenses for website fingerprinting." USENIX 2014.

Cai, Xiang, et al. "A systematic approach to developing and evaluating website fingerprinting defenses." CCS 2014.

- Very good performance
- Vulnerability approximators

Cherubin, Giovanni. "Bayes, not naïve: Security bounds on website fingerprinting defenses." PoPETs 2017

Rimmer, Vera, et al. "Automated website fingerprinting through deep learning." Arxiv 2017/NDSS 2018

Sirinam, Payap, et al. "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning." CCS 2018.

"Derive security bounds for any WF defense, where the bounds depend on a ***chosen feature set****."*

"...Such error can be estimated in practice, and is a lower bound for a WF adversary, for any classification algorithm he may use."

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. “Website fingerprinting in onion routing based anonymization networks.” WPES 2011.

Dyer, Kevin P., et al. “Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail.” IEEE S&P 2012.

Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014.

Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.

- Very good performance
- Vulnerability approximators

Cherubin, Giovanni. “Bayes, not naïve: Security bounds on website fingerprinting defenses.” PoPETs 2017

Rimmer, Vera, et al. “Automated website fingerprinting through deep learning.” Arxiv 2017/NDSS 2018

Yan, Junhua, et al. “Feature selection for website fingerprinting.” PoPETs 2018.

Sirinam, Payap, et al. “Deep fingerprinting: Undermining website fingerprinting defenses with deep learning.” CCS 2018.

*“By focusing on only a limited set of features, prior work does not help us understand the “extents” of **learn-ability (and vulnerability)** from TCP/IP headers...”*

“...what is the list of all TCP/IP features that are informative for website fingerprinting?”

...2016

Incremental Steps:
- Feature Engineering
- Performance

Panchenko, Andriy, et al. “Website fingerprinting in onion routing based anonymization networks.” WPES 2011.

Dyer, Kevin P., et al. “Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail.” IEEE S&P 2012.

Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014.

Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.

2017-2018

- Very good performance
- Vulnerability approximators

Cherubin, Giovanni. “Bayes, not naïve: Security bounds on website fingerprinting defenses.” PoPETs 2017

Rimmer, Vera, et al. “Automated website fingerprinting through deep learning.” Arxiv 2017/NDSS 2018

Yan, Junhua, et al. “Feature selection for website fingerprinting.” PoPETs 2018.

Sirinam, Payap, et al. “Deep fingerprinting: Undermining website fingerprinting defenses with deep learning.” CCS 2018.

2019-2020

- Additional Dimensions

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions

*“Even though the most effective WF attacks use deep learning to automatically extract features instead of manually craft them, manually crafted features are still important for **interpretable** machine learning.”*

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions

"Even though the most effective WF attacks use deep learning to automatically extract features instead of manually craft them, manually crafted features are still important for ***interpretable*** *machine learning."*

"In WF, finding and evaluating manually designed features can help in understanding why some sites may be especially vulnerable and how to design more effective and efficient defenses. We thus explore new ***timing features*** *in this work."*

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

*"We suspect that maintaining a perfect WF system is **costly** as the adversary needs to collect information about different localized versions of the webpages, user's browsing settings and update the system over time to recover from **data staleness**."*

Juarez, Marc, et al. "A critical evaluation of website fingerprinting attacks." CCS 2014.

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

"Over time, the community has revised and improved these assumptions to be more realistic, but some of the key assumptions have not been considered in the design of attacks. In this work, we examine how an adversary can engineer ***attacks that are more applicable in a real-world environment****."*

Sirinam, Payap, et al. "Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning." CSS 2019.

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

“Over time, the community has revised and improved these assumptions to be more realistic, but some of the key assumptions have not been considered in the design of attacks. In this work, we examine how an adversary can engineer attacks that are more applicable in a real-world environment.”

***Generalizability**. The WF classifier should be robust to the data mismatch issues that occur as a result **staleness** of training data...*

Sirinam, Payap, et al. “Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning.” CSS 2019.

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

"Over time, the community has revised and improved these assumptions to be more realistic, but some of the key assumptions have not been considered in the design of attacks. In this work, we examine how an adversary can engineer attacks that are more applicable in a real-world environment."

Generalizability. *The WF classifier should be robust to the data mismatch issues that occur as a result staleness of training data...*

Bootstrap time. *The total amount of time required for the attacker to produce a ready-to-use classifier...*

Sirinam, Payap, et al. "Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning." CSS 2019.

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

"Over time, the community has revised and improved these assumptions to be more realistic, but some of the key assumptions have not been considered in the design of attacks. In this work, we examine how an adversary can engineer attacks that are more applicable in a real-world environment."

***Generalizability**. The WF classifier should be robust to the data mismatch issues that occur as a result staleness of training data...*

***Bootstrap time.** The total amount of time required for the attacker to produce a ready-to-use classifier...*

***Flexibility & Transferability**. The WF classifier should enable the attacker to flexibly add new sites to the monitored set...*

Sirinam, Payap, et al. "Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning." CSS 2019.

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

The model still needs to be retrained when new classes are introduced.

S&P 2012.

Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014.

Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.

Yan, Junhua, et al. “Feature selection for website fingerprinting.” PoPETs 2018.

Sirinam, Payap, et al. “Deep fingerprinting: Undermining website fingerprinting defenses with deep learning.” CCS 2018.

Sirinam, Payap, et al. “Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning.” CSS 2019.

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

*“... lowers the **likelihood of data staleness** performance issues...”*

Sirinam, Payap, et al. “Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning.” CSS 2019.

Bhat, Sanjit, et al. “Var-CNN: A data-efficient website fingerprinting attack based on deep learning.” PoPETs 2020

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016

2017-2018

2019-2020

Incremental Steps:
- Feature Engineering
- Performance

- Very good performance
- Vulnerability approximators

- Additional Dimensions
- More Realistic Adversaries

“... lowers the ***likelihood of data staleness*** *performance issues...”*

“...allows a weaker attacker with ***fewer data collection*** *resources to successfully perform a powerful WF attack.”*

Sirinam, Payap, et al. “Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning.” CSS 2019.

Bhat, Sanjit, et al. “Var-CNN: A data-efficient website fingerprinting attack based on deep learning.” PoPETs 2020

Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

...2016	2017-2018	2019-2020
Incremental Steps: **- Feature Engineering** **- Performance**	**- Very good performance** **- Vulnerability approximators**	**- Additional Dimensions** **- More Realistic Adversaries** **- Transfer Learning**
Panchenko, Andriy, et al. “Website fingerprinting in onion routing based anonymization networks.” WPES 2011. Dyer, Kevin P., et al. “Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail.” IEEE S&P 2012. Wang, Tao, et al. “Effective attacks and provable defenses for website fingerprinting.” USENIX 2014. Cai, Xiang, et al. “A systematic approach to developing and evaluating website fingerprinting defenses.” CCS 2014.	Cherubin, Giovanni. “Bayes, not naïve: Security bounds on website fingerprinting defenses.” PoPETs 2017 Rimmer, Vera, et al. “Automated website fingerprinting through deep learning.” Arxiv 2017/NDSS 2018 Yan, Junhua, et al. “Feature selection for website fingerprinting.” PoPETs 2018. Sirinam, Payap, et al. “Deep fingerprinting: Undermining website fingerprinting defenses with deep learning.” CCS 2018.	Sirinam, Payap, et al. “Triplet Fingerprinting: More Practical and Portable Website Fingerprinting with N-shot Learning.” CSS 2019. Bhat, Sanjit, et al. “Var-CNN: A data-efficient website fingerprinting attack based on deep learning.” PoPETs 2020 Rahman, Mohammad Saidur, et al. "Tik-Tok: The utility of packet timing in website fingerprinting attacks." PoPETS 2020

Transfer Learning

“Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task.”

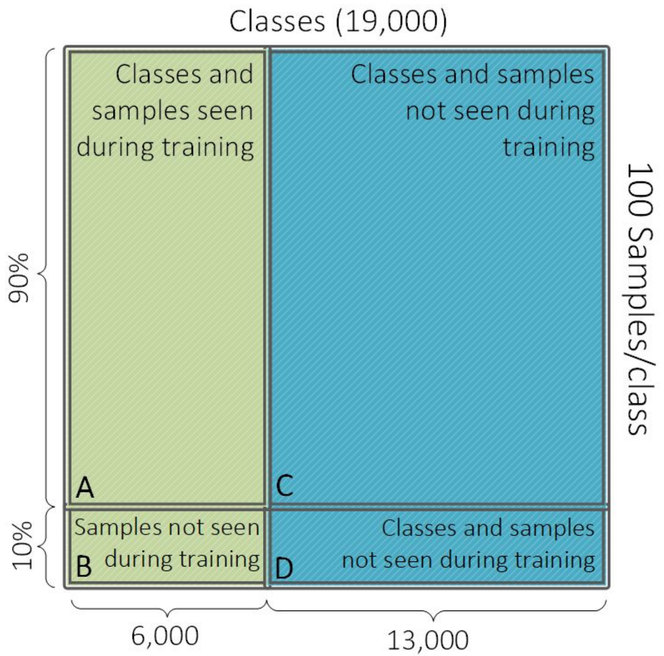
Fingerprinting models transferable across various dimensions:

- Temporal (e.g., data staleness)
- Websites/Webpages
- Location
- Protocol versions

The goal is to explore how versatile adversaries can become across those dimensions.

Wikipedia Dataset

- TLS traffic traces for Webpage fingerprinting
- 19,000 Wikipedia articles
- Each loaded 100 times
- 100 different AWS instances
- Instances spread across 5 different regions



Fingerprinting Pipeline

1. Train an embedding model
   ↪️ Maps inputs into a multidimensional space

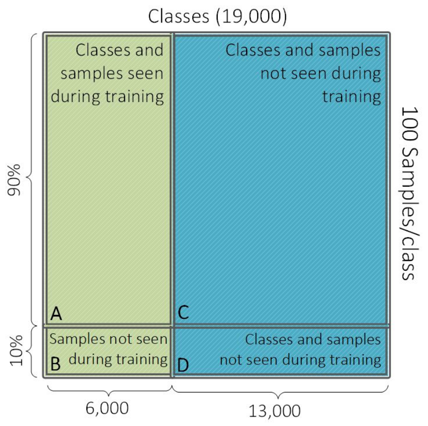
2. Gather a set of *reference* samples and embed them.

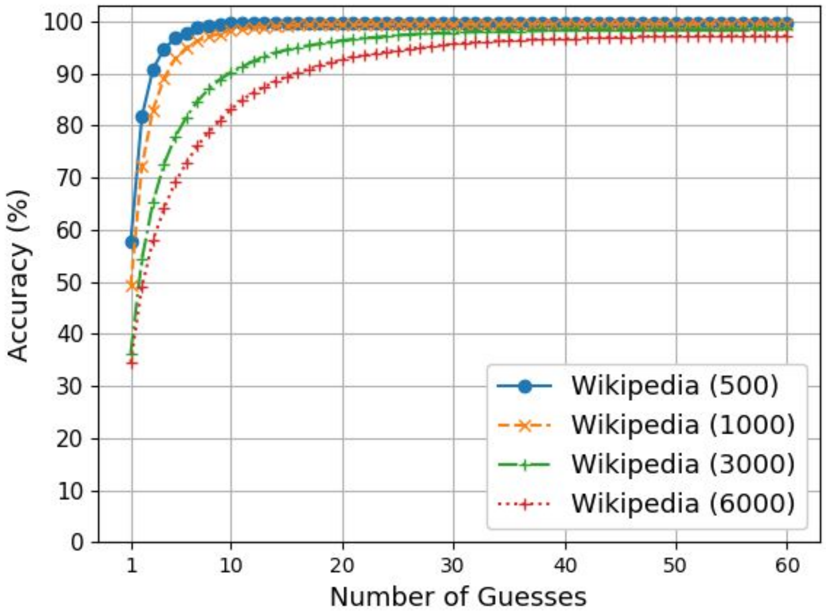
3. For each input, embed and classify based on its proximity to the reference samples.

Experiment 1

- Train on subset A (90 samples x 6,000 classes).

- Test on subset B (10 unseen samples x 6,000 classes).



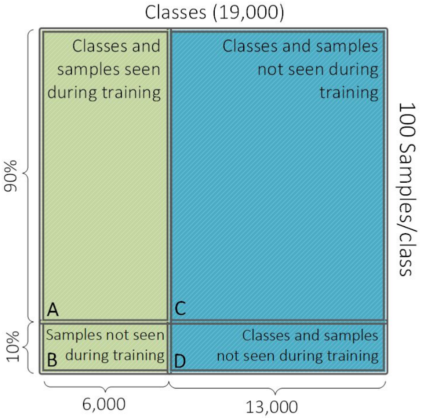


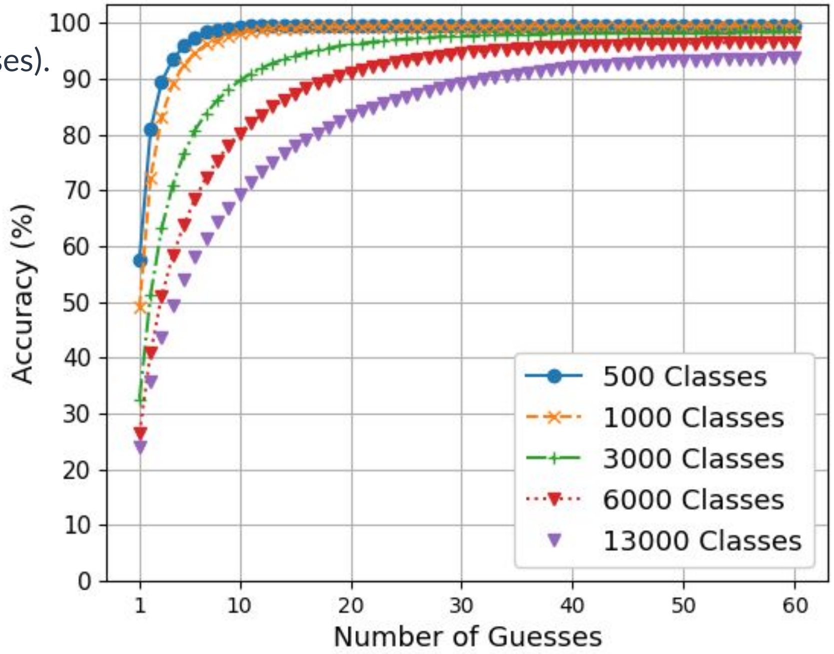
Experiment 2

- Train on subset A (90 samples x 6,000 classes).

- Test on subset D (10 unseen samples x 13,000 **different** classes).

- We use 90 samples (C) as reference and classify 10 (D).





Takeaways

Takeaways

1. The fingerprinting landscape changed rapidly due to ML/AI advancements.

Takeaways

1. The fingerprinting landscape changed rapidly due to ML/AI advancements.
2. Existing adversarial models have reached *very* high accuracy.

Takeaways

1. The fingerprinting landscape changed rapidly due to ML/AI advancements.
2. Existing adversarial models have reached *very* high accuracy.
3. New more realistic adversaries became possible.

Takeaways

1. The fingerprinting landscape changed rapidly due to ML/AI advancements.
2. Existing adversarial models have reached *very* high accuracy.
3. New more realistic adversaries became possible.
4. ML/AI has been applied mostly for fingerprinting attacks but applications in countermeasures have received less attention.

Thank you!

Questions?

Vasilios Mavroudis
University College London

Jamie Hayes
University College London