University College London

Dept. Computer Science

Automatic Feature Selection for Website Fingerprinting

Project Plan

Axel Goetz

Supervisor:
Dr. George Danezis
Jamie Hayes

November 16, 2016

1 Problem Statement

Anonymity networks like Tor use what is called onion routing where each layer in the onion represent a new layer of encryption. This allows Tor users to freely browse the web without an ISP, government, or anyone else that might be able to sniff the traffic before the first Tor node to see which websites or services the user is accessing. However even with various layers of encryption, an attacker might still be able to infer which web page a client is browsing by performing a *website fingerprinting attack*. The attack often uses machine learning to identify several trends in the network traffic such as the number of packets per second, their size, etc. But most of these attacks rely on a trail and error process of picking the features. Hence, there is no guarantee that the features used are the most appropriate ones or even any good at all. Therefore this project will analyse the use deep learning techniques such as stacked autoencoders to automatically identify features and test their effectiveness compared to the hand-picked ones in various different models.

2 Aims and Objectives

The aims and objectives for this project are as follows:

1. **Aim:** Critically review the effectiveness of current website fingerprinting attacks.
   Objectives: 1. Analyse various models that are currently used in fingerprinting attacks. 2. See if there are any flaws in the reasoning or the experimentation. 3. Examine how would a small percentage of false positives impacts a potential attack. 4. Analyse how the rapid changing nature of some web pages would impact the attack.

2. **Aim:** Generate features automatically using deep learning techniques for a website fingerprinting attack.
   Objectives: 1. Examine different deep learning feature selection methods such as stack autoencoders. 2. Pick the most appropriate method for a website fingerprinting attack. 3. Collect the necessary data to train the feature selection method. This includes a dataset that is collected over a short period of time (days) and another one that would be collected over an extended period of time (weeks). 4. Extract a set of features using this data. 5. Compare these features to existing hand-picked ones.

3. **Aim:** Train existing models with the automatically generated features and test their effectiveness compared to hand-picked ones.
   Objectives: 1. Identify various models that could be used to test the new features with. 2. Implement the models. 3. Identify various sets of hand-picked features to compare the automatically generated features with. 4. Train those models using both hand-crafted features and the generated features. 5. Compare the effectiveness of the generated features to hand-picked ones in those models in a closed-world environment. 6. Compare their effectiveness in an open-world scenario. 7. Analyse if the feature selection technique can find persistent features that are spread across a period of time and study if this helps with the classification over time. 8. Compare the effectiness of the automatically generated features when a user uses various common defeneses against website fingerprinting like camouflage. 9. Test the attack on tor hidden services as opposed to websites. 10. Investigate an appropriate technique for evaluating the result. 11. Analyse which features tend to be the most informative (highest entropy).

3 Deliverables

The project aims to produce the following deliverables:

1. Summary of website fingerprinting attacks. This includes any related work and an analysis how effective a fingerprinting attack could be (see Aim 1).

2. An analysis of the most appropriate automatic feature generation model.

3. A dataset to train both the feature generation model and the models used for the website fingerprinting attack.

4. Fully documented source code for generating the features and the models used for the attack.

5. A strategy for testing the models.

6. An analysis of using the generated features compared to hand-picked one using diferent models. This includes how the feature generation process might be able to identify persistent features that are spread across a period of time.

4 Work Plan

Project start to end October (4 Weeks)
Research current website fingerprinting attacks.
Research various method for automatic feature selection.

Mid-October to mid-November (4 Weeks)
Refine aims and objectives.
Further research into using automatic feature selection for website fingerprinting attacks.

November (4 weeks)
Collect necessary data.
Initial experimentation with feature selection.

End November to mid-January (8 weeks)
Implement feature selection.
Implementation of various models used for attacks.
Research on how to evaluate the effectiveness of a model.
Work on the Interim Report.

Mid-January to mid-February (4 weeks)
Perform tests and evaluate the performance.

Mid-February to end of March (6 weeks)
Work on Final Report.