University College London

Dept. Computer Science

Automatic Feature Selection for Website Fingerprinting

Project Plan

Axel Goetz

Supervisors:
Dr. George Danezis
Jamie Hayes

January 24, 2017

1 Current Progress

We have slightly changed how some requirements were structured in the work plan. Rather than first collecting a large dataset of website traces over a long period of time, we decided to work with existing datasets to perform some experiments and to collect that data later in the project.

Given the nature of the challenge, a majority of the time will be spend on researching models that perform automatic feature selection. After careful examination, there are only a couple deep learning methods that seem to be appropriate. These include a *Stacked Autoencoder*, a *RNN Encoder Decoder* or a *Bidirectional RNN Encoder*. In addition to simply training these models on the data, we might perform *denoising* on the them, which essentially means learning the models to distinguish uncorrupted data from corrupted data. This final step allows us to fine tune the encoders to get a consistent performance even if some of the data is noisy.

Although we have started experimenting with some of these models, we have not managed to implement all of them. However we are still on schedule for doing so.

Finally, we have also selected a set of previously successful website fingerprinting attacks. An environment has been set up, majority of these models have been implemented and the infrastructure is in place to extract a set of popular features from the raw data. This will allow us to quickly perform a performance comparison with the manually engineered features and the automatically selected ones.

2 Remaining Work

First, we will need to start the data collection process to collect data over a longer period of time. The reason why we perform this process ourselves is because there are simply no such datasets out there. This might be because the content of websites often changes rapidly, affecting the outcome of several attacks. Therefore having this data will allow us to perform an analysis of how certain attacks perform, given that the content of some web pages might have changed, which is a more realistic scenario anyway.

Next, the usefulness of some models such as the *Stacked Autoencoder* still need to be examined since it assumes a fixed-length vector as an input. Perhaps we will need to experiment with transforming the data to *wavelet coefficients* or a similar technique.

We will also need to fully implement those models and decide on the criteria that we will use for comparing the performance. Using these criteria we will then have to perform a thorough analysis of how the automatically engineered features compared to the hand-picked ones.

Finally, after all of the above has been completed, we will focus on finished writing the final report.