Side-Channel Attack: Website Fingerprinting Report

Student ID: 2005102
June 20, 2025

1 Introduction

This report outlines the findings of a side-channel attack assignment focused on website fingerprinting through the Sweep Counting Attack technique. The objective is to determine which website a user is accessing by analyzing subtle cache access patterns, without directly monitoring the screen or network traffic.

2 Latency Results Observation

N	**Median Access Latency (ms)**
1	0.00
10	0.00
100	0.00
1,000	0.00
10,000	0.00
100,000	0.10
1,000,000	1.60
10,000,000	8.60

Table 1: Access latency across increasing cache line reads

From the data, we infer that browser timing resolution is not sufficient for extremely small memory operations, but once the access count crosses 10^5, significant latency is measurable.Small memory reads are too fast for browsers

to time accurately, but large reads take enough time to be measured. So, to detect cache behavior we need to work with big enough memory operations

3 Trace Collection with Sweep Counting

3.1 Setup

- Cache line size: 64 bytes
- Last Level Cache (L3): 8 MB

3.2 Heatmap Visualizations

The sweep count data was used to generate heatmaps. Below are the images for different browsing states.

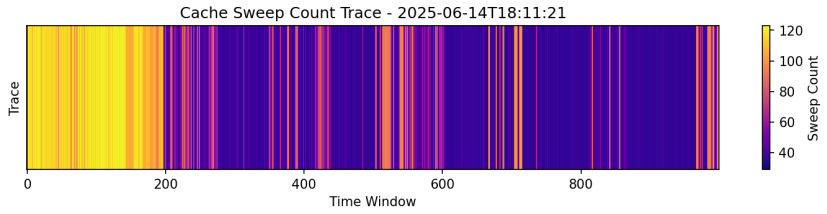

Figure 1: Heatmap for cse.buet.ac.bd/moodle

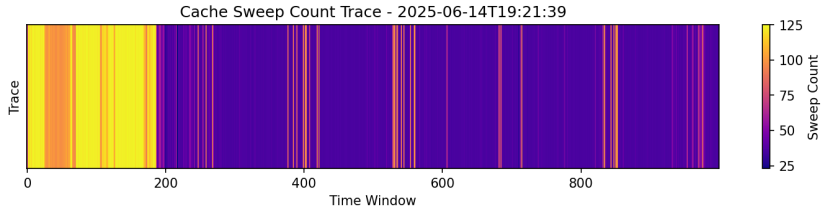

Figure 2: Heatmap for google.com

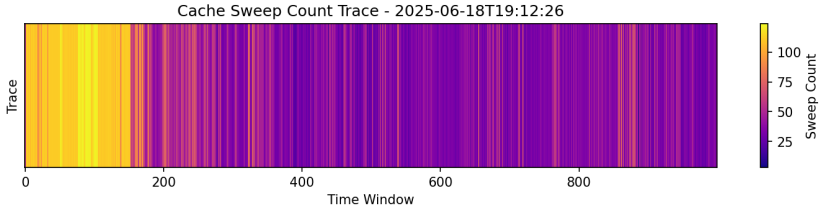

Figure 3: Heatmap for prothomalo.com

These differences in cache usage lead to unique interference patterns that are captured by our sweep function, creating distinguishable traces.

3.3 Choosing a Suitable Value for P

The value of P (duration of one sweep window) must:

- Be small enough to collect many samples in the 10-second duration for better resolution.
- Be large enough so that the time taken to perform a sweep is actually measurable using performance.now().

Based on Table-1,value of P is chosen 10

4 Automated Data Collection

Using Selenium automation, we collected trace data for multiple websites with user-simulated scrolling behavior. Each trace was stored in a SQLite database for persistence.

4.1 Websites and Dataset

Traces were collected for:

- `https://cse.buet.ac.bd/moodle/`
- `https://google.com`
- `https://prothomalo.com`

4.2 Personal Data Collection

Each site had 1100 traces (a total of 3300 traces) collected by Selenium WebDriver for Google Chrome. These were used to train and test a simple and complex model (provided by template).

4.3 Combined Data Collection

Each site had 17,000 traces (a total of 51,000 traces) collected in total from different environments. For improved generalizability, trace data from multiple machines were merged and normalized per-user, ensuring that range differences between systems did not bias the model.

5 Machine Learning for Website Classification

To evaluate the effectiveness of cache-based side-channel website fingerprinting, two convolutional neural network architectures—simple_cnn and complex_cnn—were trained and tested using both personal trace data (3300 samples) and the combined dataset (51,000 samples).

Results from Personal Trace Data (3300 samples)

Simple CNN:

- Accuracy: 81.82%
- Best Performing Site: `https://prothomalo.com` with 0.95 F1-score

Classification Report:

Website	Precision	Recall	F1-score	Support
`https://cse.buet.ac.bd/moodle/`	0.76	0.71	0.74	220
`https://google.com`	0.75	0.80	0.77	220
`https://prothomalo.com`	0.95	0.95	0.95	220

Table 2: FingerprintClassifier results

Calculations:

- Accuracy: 0.82
- Macro Average: 0.82
- Weighted Average: 0.82

Complex CNN:

- Accuracy: 84.85%
- Best Performing Site: `https://prothomalo.com` with 0.95 F1-score

Classification Report:

Website	Precision	Recall	F1-score	Support
`https://cse.buet.ac.bd/moodle/`	0.80	0.78	0.79	220
`https://google.com`	0.81	0.78	0.80	220
`https://prothomalo.com`	0.92	0.98	0.95	220

Table 3: ComplexFingerprintClassifier results

Calculations:

- Accuracy: 0.85
- Macro Average: 0.85
- Weighted Average: 0.85

Results from Combined Dataset (51,000 samples)

Simple CNN:

- Accuracy: 75.03%
- Best Performing Site: `https://prothomalo.com` with 0.86 F1-score

Classification Report:

Website	Precision	Recall	F1-score	Support
https://cse.buet.ac.bd/moodle/	0.77	0.63	0.69	3400
https://google.com	0.66	0.74	0.70	3400
https://prothomalo.com	0.83	0.88	0.86	3400

Table 4: FingerprintClassifier results

Calculations:

- Accuracy: 0.75
- Macro Average: 0.75
- Weighted Average: 0.75

Complex CNN:

- Accuracy: 80.94%
- Improved Stability Across Sites: Higher precision and recall across all three websites. Indicates better generalization over noisy, heterogeneous data.

Classification Report:

Website	Precision	Recall	F1-score	Support
https://cse.buet.ac.bd/moodle/	0.78	0.77	0.78	3400
https://google.com	0.79	0.74	0.76	3400
https://prothomalo.com	0.85	0.91	0.88	3400

Table 5: ComplexFingerprintClassifier results

Calculations:

- Accuracy: 0.81
- Macro Average: 0.81
- Weighted Average: 0.81