Project Proposal

CSE 508: Network Security

1 Group

- Divyansh Upreti (SBU ID: 112026646)
- Prabuddha Kumar (SBU ID: 112076553)
- Rohit Aich (SBU ID: 112126618)

2 Title

Winking Router: A study on side channel attacks for website identification through routers

3 Problem

Side channel attacks provide an effective means for gauging usage patterns over a network. These can reveal information like the websites accessed over a network, time when machines are active etc. Our project aims to address the feasibility of a side channel attack by observing the router blinking pattern to estimate the website being accessed. This is a potential threat because modern day smartphones have high frame-rate cameras that can be used to observe the blink pattern of a router. Through the course of the project, we will try and model the signature pattern of each website and use this model to predict the website being accessed.

4 Context

There has been a lot of related work to analyze the threat side channel attacks pose for various devices in a network. The study done by *Qing Yang et al* [1] shows that public USB charging stations pose a significant privacy risk to smartphone users even when no data communication is possible between the station and the user's mobile device. Their results show that the attack is highly successful and they were able to achieve over 90% webpage identification accuracy. This work is really appreciable as it is the first to study the side channel attack on smartphones under phone constraints and it demonstrates that websites can be correctly identified within a short timespan of two to six seconds. USB charging stations are becoming widespread around the world, and modifications required to implement these attacks can be easily concealed and therefore can go unnoticed by users. The paper uses machine learning algorithms to identify the webpage the user visited. The researchers take care of smartphones factors such as user interaction with touchscreen, WiFi and LTE connectivity, training and testing device mismatch, type of connection (HTTP or HTTPS) and relative location of the host serving the webpages to the smartphone.

Xun Gong et al [2] analyze the web traffic. They have shown that traffic patterns leaked through side channels can be used to gain sensitive information. Attackers can find out which website, or webpage a user is accessing simply by monitoring the packet size distribution. The attacker sends frequent probs and measures the sizes, timings and counts of packets arriving. If the training and test data collected is from the same location, large fraction of sites can be accurately detected. They study on how traffic analysis can be a serious threat to Internet privacy as it can be carried out remotely, without access to the analyzed traffic, thus greatly increasing the scope of attack.

Edward W. Felten et al [3] studied the attacks that allow a malicious web site to determine whether or not the user has recently visited some other, unrelated web page. This malicious page can determine this by measuring the time the user's browser requires to perform certain operations. As browsers perform various forms of caching, the time required for operations depends on user's browsing history. By measuring the time required to access an element, malicious program can tell if the element is in a nearby cache and thus learn that element has been accessed recently. This is a dangerous side attack to conceal user's browsing history and figure out his general browsing patterns.

5 Approach

We plan to start this experiment in an controlled environment. We could use an exclusive browser, running on a dedicated subnet, with only one single computer connected with it. We could use a Raspberry Pi board for this as well.

We plan to have a parallel approach for this experiment. Firstly, we will have to get a sensor that will capture the blinks of the router LED, as each packet arrives. We assume that this sensor capture the data (count of blinks) correctly. We now plan to device an experimental way to empirically study and conduct our research. We plan to do this in three rounds of laboratory tests. We will consider top 50 Alexa global sites for our research.

- At first, we hit the sites, after omitting any external disturbances, i.e., we just load the web-pages, but block any crawlers, scripts, malwares etc. that might come with it. We will be using a few readily available browser plugins for this which detect and stops unintended scripts from running on a local browser.
- Next, we plan to hit the websites in an controlled environment. In this iteration, we plan to clear the browser cookies and cache after each hit, so that we can record the actual page loading event, and the browser would be unable to cache a few packets.
- Lastly, we plan to hit the websites in an uncontrolled environment, in presence of disturbances. This run would record the performances when the pages are hit in real world scenarios, with cache and cookies being stored, and hidden browser scripts being downloaded as well. We could even plan to scale the research up and connect a more few machines to the router to understand whether the router could cache a few packets from the web-pages.

During all the above iterations, our goal would be to develop an Machine Learning training set to fingerprint the websites. We plan to plot the packet arrival rate over time and bandwidth, and make an optimal pattern for each website, optimized over thousands of hits. Also, we plan to find the average response times of these websites, and the standard deviations. Combined with the packet arrival pattern, this data could also help us understand which web-page is being hit.

We plan to write pieces of code which will automatically complete steps 1 and 2 for us, and record the data.

Once our training set is developed, we can then find the arriving packets from the number of blinks, and then plot it over the mean response time to generate a pattern, same as our training sets. We will then use those training data and try to predict the websites from these test sets (patterns). Based on the accuracy of the prediction, we would be able to understand whether, and to what degree, there is a correlation between router blinks and web-pages which are being hit.

6 Evaluation

The objective of the project is to use machine learning to identify signature patterns of packet transfer for websites using the blinking lights of a router. In order to do this, we will need data of the following form:

- We will need to record the the timestamp of each packet received for an ideal sample. Here, an ideal sample refers to the sample that comes from having a stable network speed and only the website's packets passing through the router
- We repeat this experiment several times to get a large dataset. After having a large sample, we can model the curve for the ideal data for each website.
- Using the generated model, we can attempt to predict the website on non-ideal samples of the following kind:
- Fluctuating network speeds: The speed of the network is unstable and can cause the timestamp to proportionally scale in or out.

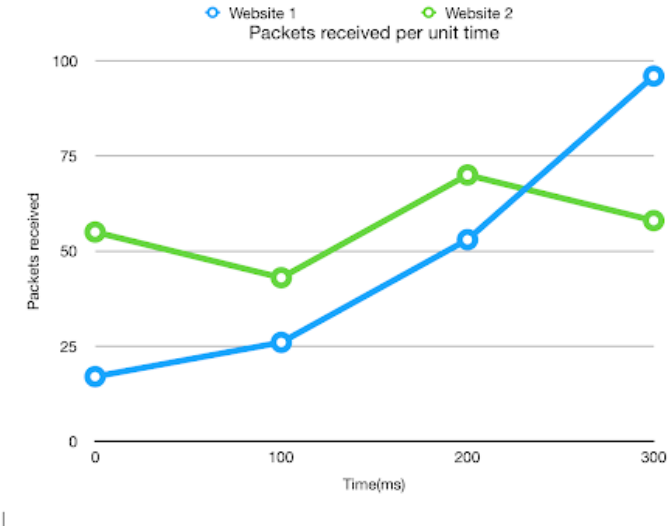

Figure 1: Packets received per unit time for 2 websites

- Router being used by other applications/devices: The router is relaying traffic from other applications or devices. This would pollute the data and make the detection task significantly harder.
- We can attempt to train the model to work with fluctuating network speeds and with a busy router to test the limits of the model.

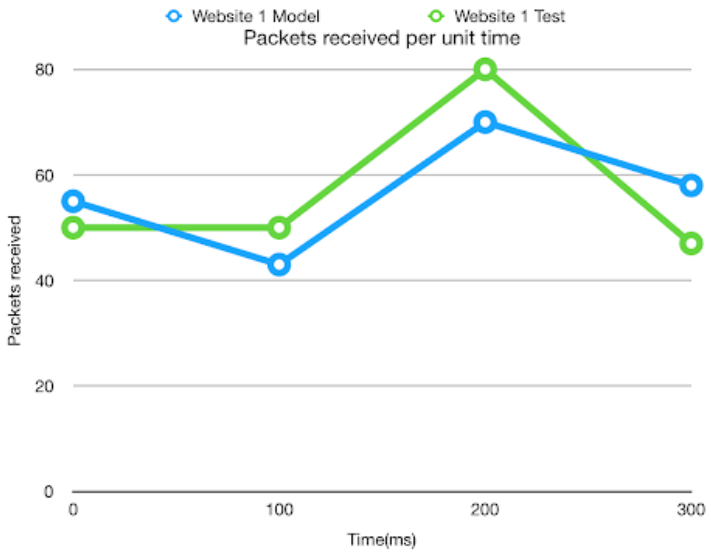

Figure 2: Model estimate where Website 1 Test is the plot during testing. Website 1 Model is the modeled chart

- Through this way, we test the prediction and attempt to improve it and make it more robust across different conditions of operation.
- To evaluate, we train over a large sample, and test our model to observe the accuracy of prediction for the following cases:
  - Ideal case
  - Fluctuating network speeds
  - Busy Router

7 Scope

The scope of the work includes the following:

- Exploring whether there are any relation between the arriving number of network packets and the web-pages being hit, and using light as a side channel.
- Trying to fingerprint websites based on their specific packet arrival pattern over time and bandwidth. We are planning to approximate curves for individual websites and develop a robust training set.
- This work has huge potential in the field of side-channel attacks with light emissions. Till date we have had researches where rate of packet arrival or the time for a page to load could be mapped with specific web-pages up to a certain accuracy. In this study, we are extending those researches to a more pragmatic problem, by collaborating packet arrival rate and response-time data to deduce specifically which web-page was hit. Furthermore, the linear relation between router LED blinks and network packet arrival rates gives us a unique edge to map the router blinking patterns to specific websites. We plan to start with pristine research data (without noise), and later scale the research to real-world conditions.
- This study also uses common smartphone cameras as LED sensors. This inclusion is extremely crucial, as smartphones has turned to a public gadget now. We intend to show that, by simply pointing smartphone cameras to router LED-s, we can predict about the web-pages that are being accessed. We plan to coin a probability value by which this can be deterministically predicted.

- The final goal is to pave the way for a future study to see how far light-emission based side channel attacks are efficient to predict a browser's online activity.
- The scope does not include the study of similar attacks on IoT devices. Our research is strictly based on PC-s.

8 Approximate Timeline of Project

Following is the approximate timeline of this project.

Date of Completion	Objective
20/08/2018	1. Hardware system ready for study.
	2. A script to automate the experiment installed on machine that will clear cache and cookies and terminate the background processes periodically.
01/11/2018	1. Prepare the data set for training and testing.
22/11/2018	1. Train the data into machine, observe results, draw inferences, plot graphs and if possible scale the problem.
06/12/2018	1. Final analysis of results and preparation of final report.