CS6570: SSE Project Abstract

Rahul Kejriwal, CS14B023

Project: “Searching through encrypted traffic to identify websites”

Abstract:

Encrypted communication can prevent adversaries from discerning information about user activity by sniffing network logs obtained from tools like Wireshark, however, vanilla use of encryption can still allow the possibility of information leakage. This project aims to perform a side-channel attack on encrypted communication between users and servers to identify the domain/website the user is accessing.

The rate of transmission, packet size distribution, no. of packets transmitted etc. are features of communication that differ slightly among different domains and this information can be used to fingerprint the website and allow a potential attacker to analyze a user’s network logs to identify some websites which the user had accessed.

Various defenses like application-level defenses HTTPOS and randomized pipelining over Tor have been proposed to prevent such information leakage. The project will also study their effectiveness and whether they can be circumvented.

Expected Learnings:

The project will aim to discover most revealing features of encrypted traffic that leak domain information and evaluate effectiveness of currently proposed defenses.

Expected Demonstrable Results:

The project should result in a final model for website identification for top 100 websites from a user’s network logs containing encrypted traffic.

Input:	Network Log of user activity from WireShark
Output:	List of detected websites (from the top 100 on which model is trained)

The model should work across multiple systems (should not be tuned to 1 system) and multiple users (should not be tuned to 1 user’s activity), i.e., the model should generalize well.

Approach:

The project will try to study and adapt the work done by Xiang Cai et al. [1] and Herrmann et al. [2] in order to attempt the attack and study effectiveness of defense measures used by real-world websites. A nice summary of the work done in this regard has also been given by Mike Perry [3].

Timeline:

Week 1: (16th - 22nd Sep)	**Week 2: (23rd - 29th Sep)**	**Week 3: (16th - 22nd Sep)**	**Week 4: (30th Sep - 6th Oct)**	**Week 5: (7th - 13th Oct)**	**Week 6: (14th - 20th Oct)**	**Week 7: (21st - 25th Oct)**
Literary Review & Understanding Problem Space		Data Collection/ Aggregation		Model Training, Testing & Refinement Evaluation of Defence strategies (if time permits)		Documentation & Presentation

References:

[1] Cai, Xiang, et al. "Touching from a distance: Website fingerprinting attacks and defenses." *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012.

[2] Herrmann, Dominik, Rolf Wendolsky, and Hannes Federrath. "Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier." *Proceedings of the 2009 ACM workshop on Cloud computing security*. ACM, 2009.

[3] Perry, Mike. “A Critique of Website Traffic Fingerprinting Attacks.” Tor Blog, Tor Blog, 7 Nov. 2013, blog.torproject.org/critique-website-traffic-fingerprinting-attacks.