Master's Project Pre-Proposal: (Profiling of malicious web application users)

Kyle Marcotte

February 10, 2011

1 Problem Description

This master's project will demonstrate a methodology for profiling malicious web application users by their evolving browser fingerprint, location, and ip address. Browser fingerprints can be defined by the information that browsers provide to webservers when they make a connection. Furthermore, this project will demonstrate the usage competitive Learning Vector Quantization networks to maintain and adjust these intruder profiles as the fingerprint and other information associated with the user evolves.

Project Goal Development and testing the technique of detecting reentry of malicious users based on their unique browser fingerprint. Comparisons will be made between a competitive Learning Vector Quantization network, a modular MultiLayer Perceptron Classifier, and a k-nearest neighbor classifier ensemble.

2 Importance of Research

Conventional intrusion detection techniques are very reactionary in nature. The current direction in security research points towards methods that involve building and maintaining profiles of users either for the purpose of preventing masquerade attacks or for monitoring suspicious activity of likely attackers. Likely attackers can be identified through the use of honeypots or honeynets. [2] The challenge that these methods present is that the information from which these profiles are built is constantly changing from session to session. In order to maintain useful profiles, whatever profiler is used must be robust enough to deal with the constantly changing user data.

3 Related Work

Eckersley performed detailed research into the degree to which modern web browsers are subject to fingerprinting via the version and configuration information that they will transmit to websites upon request. He described the type of information that can be collected and performed an investigation towards understanding the entropy of the system. [1] Kwan et. al. have proposed a honeypot/honeynet approach as a method of identifying suspicious user activity on the web. Some methodology of identifying suspicous users from the population of all users is necessary for narrowing the size of the population to be profiled. [2] Marin et. al. have investigated the applicability of Learning Vector Quantization (LVQ) as a technique for profile creation and intrusion detection. LVQ is a competetive learning technique that clusters data into disjoint polytopes to approximate the decision surfaces of a Bayesian classifier. [3]

4 Methodology

A competitive Learning Vector Quantization network will be implemented to profile a subset of users and adjust to the user's evolving browser fingerprint. In addition to the LVQ network, a k-nearest neighbor classifier and a modular MLP classifier will be implemented to provide a baseline for comparisons. Each classifier will be constrained to use an identical input format and produce an equal number of classes in order to preserve the validity of the comparisons. A web page will be developed and placed on the CS server to collected fingerprinting data from visiting users. Each classifier will be trained on a training subset of the collected data. The remaining data will be used to test the classifiers. A sample of results will be collected from each classifier and compared using a Wilcoxon signed-rank test. For this test the performance metric used will be the number of correct classifications.

5 Potential Outcomes

LVQ outperforms MLP A recommendation will be made for further investigation of competetive learning techniques for the problem of intruder profiling.

MLP outperforms LVQ A recommendation will be made for further investigation of supervised perceptron network techniques for the problem of intruder profiling.

References

[1] Peter Eckersley. How unique is your web browser? Technical report, Electronig Frontier Foundation, 2009.

[2] Leonard Kwan, Pradeep Ray, and Greg Stephens. Towards a methodology for profiling cyber criminals. In *Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, HICSS '08, pages 264–, Washington, DC, USA, 2008. IEEE Computer Society.

[3] J. Marin, D. Ragsdale, and J. Sirdu. A hybrid approach to the profile creation and intrusion detection. In *DARPA Information Survivability Conference Exposition II, 2001. DISCEX '01. Proceedings*, 2001.