Detecting Phishing Websites based on Machine Learning Classifier

John Onaya

(B135/10932/2014)

Supervisor: Dr. Boaz Too

Abstract

There are number of users who purchase products online and make payment through various websites. There are multiple websites who ask user to provide sensitive data such as username, password or credit card details etc. often for malicious reasons. This type of websites is known as phishing website. In order to detect and predict phishing website, we proposed an intelligent, flexible and effective system that is based on using classification Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to classify their legitimacy. The phishing website can be detected based on some important characteristics like URL and Domain Identity, security and encryption criteria in the final phishing detection rate This project proposes a phishing detection plugin for targeting chrome browser that can detect and warn the user about phishing web sites in real-time using random forest classifier. One common approach is to make the classification in a server and then let the plugin to request the server for result. Unlike the old approach, this project aims to run the classification in the browser itself. The classifying in the client side browser has advantages like, better privacy, detection is independent of network latency. This project is implemented in Javascript for it to run as a browser plugin. Since javascript doesn't have much ML libraries support and considering the processing power of the client machines, the approach needs to be made lightweight.