Contribution ID: **10** Type: **Poster**

Enhancing Phishing Detection on Websites: A Hybrid Approach Combining Signature-Based Detection and Traditional Machine Learning Methods - Abstract

Thursday 24 April 2025 10:45 (30 minutes)

With the increasing popularity of web services and the significant amount of time users spend online, cybercriminals are increasingly targeting this space with sophisticated malicious techniques. Phishing, one of the most prevalent cybersecurity threats, poses a substantial risk to both individuals and organizations. These attacks are typically executed via deceptive emails containing fraudulent URLs that redirect users to malicious websites designed to steal sensitive information. The consequences of phishing can be severe, including financial loss, identity theft, and malware infections. To mitigate phishing threats, implementing an effective detection framework is essential. Existing detection techniques include signature-based methods, traditional machine learning models, and deep learning approaches. However, each method has inherent limitations, making it challenging to rely solely on a single technique. We propose a hybrid phishing detection approach that integrates signature-based detection with a machine learning model to enhance both versatility and robustness against phishing attacks. Email file in EML format works as an input to our solution which is later processed to extract hyperlinks from body of the message. Those URLs are then passed to signature-based detection component and machine learning model. The first component utilizes an API wrapper for the widely used cyber-threat intelligence platform. Meanwhile, the machine learning model is trained on feature extraction techniques using the open-source phishing URL dataset. Those features are based on URL attributes (e.g. domain length, HTTPS usage, number of special characters) and HTML attributes (e.g. number of hidden fields, redirections, JavaScript references). For the research purposes, we trained three different machine learning models: Decision Tree, MLP (Multi-Layer Perceptrons network), and SVM (Support Vector Machines). All of our models achieved an impressive accuracy of 0.9988, 0.9994, and 0.9982, respectively, demonstrating its effectiveness in identifying phishing threats. The results indicate that this hybrid approach can serve as a reliable cybersecurity tool, capable of detecting both existing and emerging phishing attacks with high precision.

Authors: ROMANEK, Jakub (Warsaw University of Technology); LATAWIEC, Marcin

Co-author: Dr CHABER, Bartek

Presenters: ROMANEK, Jakub (Warsaw University of Technology); LATAWIEC, Marcin

Session Classification: Session B (Poster)