14.01.2024

Methods For Detecting Cybersecurity Attacks

Final Project - Research Proposal

Subject: Suspicious URL detection using machine learning models.

Team Members:

1. Yoad Tamar,
2. Lior Vinman,
3. Nevo Gadassi.

Paper: "Dipankar Kumar Mondal, Bikash Chandra Singh, Haibo Hu, Shivazi Biswas, Zulfikar Alom, Mohammad Abdul Azim - SeizeMaliciousURL: A novel learning approach to detect malicious URLs ".

Abstract:

1.
    a. ***The Problem It Tries to Solve***:
    The article tackles the challenge of detecting malicious URLs and presents a novel approach aimed at enhancing classification accuracy. In essence, the paper outlines machine learning methods for determining the suspicious nature of a given URL.

    b. ***Proposed Approach***:
    We can break down the proposed approach into three main steps of learning and processing.
    <u>The first stage</u> is "**Ensemble Learning**" In this stage, multiple classifiers are employed. There is a deployment of various machine learning models, where each model processes a different segment of the URL using distinct algorithms. At the conclusion of each model's processing, the model outputs the probability it assigns to the segment being valid or not. Through a soft voting mechanism, the outputs of all models are aggregated.
    <u>The second stage</u> is "**Threshold Filtering**" In this phase, a self-check is implemented on the outputted results. A threshold is applied to filter decisions, considering the absolute difference between class probabilities.
    <u>The last stage</u> is "**Final Decision**" In this segment, the ultimate output determining the validity of the URL is made. It considers the classes with the highest probabilities after filtering, and the chosen class serves as the final decision for the URL (whether it is malicious or non-malicious).

2. ***Main Results***:
    The main result of the experiments is that the proposed "$SeizeMaliciousURL$" model, utilizing a soft voting-based ensemble learning approach with a threshold value δ, outperformed other machine learning models in detecting and classifying malicious URLs. The model achieved the highest accuracy of 99.91% for dataset-I and 97.98% for

dataset-II, surpassing traditional supervised learning and ensemble approaches. The experiments also highlighted the effectiveness of the threshold filtering technique in refining decisions and improving the model's performance. Further, the F1-score metric confirmed the superiority of the proposed model in comparison to other approaches. Despite not achieving 100% accuracy, the results indicate promising advancements in malicious URL detection, suggesting the need for continued research and potential enhancements through feature engineering.

3. ***Our Workplan*:**
   In this project, we are planning to implement a machine learning model that will determine if an URL is suspicious. The model will learn from a large dataset and will collect all the characteristics that indicates if an URL is suspicious or not (such as: length, special characters, domains, prefixes, postfixes, protocol, etc.). In addition to the learning algorithms, we'll use some existing determination tools (for instance: virustotal) to increase the probability of correctly classifying a URL.
   Our model will function as a CLI tool (will be written in Python), receiving an URL as a main argument, then running it over our implemented model (which is trained to detect malicious URLs), and then verify with the other APIs (sort of ensemble learning) – at the end, the tool will output if the given URL is suspicious or not.

***Related Work*:**

Dipankar Kumar Mondal, Bikash Chandra Singh, Haibo Hu, Shivazi Biswas, Zulfikar Alom, Mohammad Abdul Azim - SeizeMaliciousURL: A novel learning approach to detect malicious URLs.

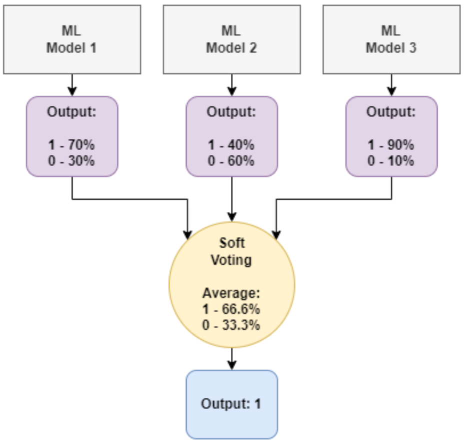
[2] Sahoo D, Liu C, Hoi SC. Malicious URL detection using machine learning: A survey. 2017, arXiv preprint arXiv:1701.07179.

[34] Haider CMR, Iqbal A, Rahman AH, Rahman MS. An ensemble learning based approach for impression fraud detection in mobile advertising. J Netw Comput Appl 2018;112:126–41.

***Dataset*:**

We'll try to use the dataset that is used in the paper, which is a .csv file of a lot of URLs with classification. Can be found here.
Moreover, we will also ask DR. Dubin for another appropriate dataset.

***Architecture*:**


ML Model 1
ML Model 2
ML Model 3
Output: 1 - 70% 0 - 30%
Output: 1 - 40% 0 - 60%
Output: 1 - 90% 0 - 10%
Soft Voting
Average: 1 - 66.6% 0 - 33.3%
Output: 1