**SEADer: A Social Engineering Attack Detection method based on Natural Language Processing and Artificial Neural Networks**

August 2019

Decision Tree: 0.918

Random Forest: 0.917

Multi-Layer Perceptron: 0.925

Dataset “*SEADER”*

After the pre-processing of the dialogues (steps 1 – 11), the classification dataset has

the following 4 labels: (1) Intent, (2) Spelling, (3) Link and (4) attack or no attack.

The SymSpellpy library (a Python port of SymSpell) was used for spellingC the Web of Trust (WOT) Application Programming Interface (API) was used to check any links and finally the SciKit library for the MLP classifier.

compound dataset is based on the 147 entries plus 600 entries from customer support-based tweets from Twitter, none of which are classified as attacks.

* Compound dataset results a higher accuracy.

**Detection of Social Engineering Attacks Through Natural Language Processing of Conversations**

2016

* Hobbs Algorithm
* Precision100%, recall 60%
* Dataset “2013 Data Breach Investigations Report” (not publicly available)

Dataset “conversations from the U.S. Supreme Court Oral Arguments” (available for purchase only)

Dataset “The penn treebank” for conversation language parsing.

Uses natural language processing techniques to

detect questions and commands.

Each extracted topic is compared against a topic blacklist to determine if the question or command is malicious.

Steps:

* Question/Command Detection
* Topic Extraction

Use the Tregex tool to match patterns in parse trees.

**Catch me, Yes we can! - Pwning Social Engineers using Natural Language Processing Techniques in Real-Time**

2018

**Dataset:** *multiple*

Information gathered might include explicitly secure information such as a credit card number or seemingly innocuous information which can support a larger attack, such as the name of a coworker.

In order for the attacker to achieve his goal, the attacker must perform one of the following detectable actions.

* **Ask a question** whose answer is private.
* **Issue a command** to perform a forbidden operation.

Uses question answering technology to provide the privacy status of the questions answer.

Evaluates commands by summarizing their meaning as a combination of the main verb and the object(s) of that verb in the sentence. *“Please reset the router” would be summarized by the verb-object pair (reset, router).*

*verb-object pair will be compared to a blacklist.*

|  |  |  |
| --- | --- | --- |
| **Database** | **URL** | **Size** |
| Scamdex | <http://www.scamdex.com> | 56555 |
| Scamwarners | <http://www.scamwarners.com> | 43241 |
| Scamalot | <http://scamalot.com> | 18149 |
| Antifraudintl | <http://antifraudintl.com> | 69209 |
| **Total** |  | **187154** |

This paper assumes that anyone communicates with the user is a stranger and doesn’t have any type of authentication to ask for any level of private data.

SOCIAL ENGINEERING ATTACK DATA

<https://www.kaggle.com/wcukierski/enron-email-dataset> kaggle (1.43 GB)

Form questions are the items in a questionnaire where each individual item is not a question, but the context tells the listener that the item should be treated as a question and an appropriate answer is expected.

Uses **Bidirectional recurrent neural network** to separate text into sentences.

extract information about each sentence by using the Stanford Parser to generate a syntactic parse using a probabilistic context-free grammar.

Uses PARALEX question answering approach for question analysis.

**AN AUTOMATED SYSTEM FOR DETECTION OF SOCIAL ENGINEERING PHISHING ATTACKS USING MACHINE LEARNING**

July 2020

**Dataset:** extracted from Github and PhishTank.

classification of phishing and legitimate URL with 96% **accuracy**.

Using Logistic Regression to classify the URLs.

**An Analysis of Various Social Engineering Attack in Social Network using Machine Learning Algorithm**

October 2020

**Dataset:** Mendeley Data contains 48 features extracted from 5000 phishing webpages and 5000 legitimate webpages.

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Accuracy without  modifying | Accuracy with  modifying | Change % |
| J48 | 98.84% | 99.01% | 0.17% |

**Algorithm:** Decision tree J48 (for URL attacks)

* This proposed technique could classify the data as either normal or abnormal.