Rule-Based Email Phishing Detection System

Introduction

Phishing is a common cyber threat that involves sending fraudulent emails with the intention of tricking recipients into revealing personal information or carrying out destructive acts. These emails frequently pose as official correspondence from reliable sources. Developing automated systems that can reliably identify phishing emails is essential to reducing the dangers associated with phishing assaults. This study describes a phishing detection system that looks at sender information, email content, and attached URLs or files, among other things, to identify phishing emails using rule-based procedures.

Implemented Rules

The phishing detection system classifies emails as legitimate or phishing based on a number of rules. These guidelines are intended to address common elements found in phishing emails, including grammatically incorrect sentences, urgent language, mismatched URLs, harmful attachments, and suspicious domain names. A brief summary of each rule is provided below:

1. **Domain Check:**

   Phishing emails often come from domains that resemble legitimate ones but contain subtle differences (e.g., "amaz0n.com" instead of "amazon.com"). The system checks the sender's domain for suspicious characteristics such as unusual top-level domains (TLDs) or domain names.

2. **Urgent Language Detection:**

   Phishing emails frequently use urgent or threatening language to create a sense of panic or urgency, compelling the recipient to act quickly without scrutinizing the email. This rule analyzes the subject and body of the email for keywords indicative of urgency (e.g., "urgent," "immediate action required").

3. **Mismatched URLs:**

   A common phishing tactic is to display a legitimate-looking URL but have it redirect to a malicious site. The system checks for such discrepancies.

4. **Malicious Attachments Detection:**

   Phishing emails may contain attachments that can execute malicious code or install malware when opened. This rule examines the types of attachments and flags those that could potentially harm the user.

5. **Grammar and Spelling Errors:**

   Many phishing emails contain spelling mistakes or grammatical errors, which are less common in legitimate business communications. The system checks the email content for such errors, adding to the overall phishing score if found.

Evaluation Results

To evaluate the effectiveness of the phishing detection system, a labeled dataset of emails, consisting of both genuine and phishing samples, was used. The system was tested on this dataset, and the following results were obtained:

- True Positives (TP): 6
- True Negatives (TN): 7
- False Positives (FP): 2
- False Negatives (FN): 2

Based on these counts, the performance metrics are as follows:

- **Accuracy:** 0.76

  Accuracy indicates the overall effectiveness of the system in correctly classifying emails. The system achieved an accuracy of 76%, showing a reasonable level of correctness in its classifications.

- **True Positive Rate (TPR):** 0.75

  The True Positive Rate, also known as Recall or Sensitivity, is 75%. This means that 75% of the actual phishing emails were correctly detected by the system.

- **False Positive Rate (FPR):** 0.22

  The False Positive Rate is 22%, indicating that 22% of legitimate emails were incorrectly flagged as phishing.

Suggestions for Improvement

While the phishing detection system performs reasonably well, there are several ways to enhance its effectiveness:

- Incorporate Machine Learning:

  Utilize machine learning models that can learn from data and dynamically improve phishing detection by understanding new patterns and evolving tactics.

- Enhance Natural Language Processing (NLP):

Improve the urgent language detection rule by using advanced NLP techniques, such as sentiment analysis and intent recognition, to better understand the context and tone of the email content.

- Expand Dataset:

  Increase the size and diversity of the dataset used for training and testing. More data will allow the model to generalize better and reduce overfitting.

- Multi-Factor Scoring System:

  Implement a more sophisticated scoring mechanism that considers multiple factors and adjusts weights dynamically based on the email's context and content.

- Behavioral Analysis:

  Include behavior-based analysis by tracking user behavior related to emails, such as clicking links or downloading attachments, to identify suspicious activities.