Ryan Yoak

Dr. Delozier

Information Security

April 27, 2020

Phishing Emails and PhyFinder

In the modern world there are still many ways malicious actors attempt to gain the sensitive information of those most vulnerable in society. One of those way is the phishing email, a message sent in an attempt to gain information about the target that can be used to possibly steal their identity, gain access to accounts, or even steal money. In fact, in 2019 $57,000,000 was lost to phishing attacks in the United States alone, and thousands of attacks are launched every day (USGOV). While those who are more confident in technology and phishing attacks are certainly susceptible to falling for a phishing attack, these attacks normally target those more gullible in our society. In fact, phishing emails often include implausible scenarios and incorrect grammar and spelling in order to make obvious to most that the email is malicious in nature to most, only leaving those already gullible or ill-informed enough to be likely to fall for a phishing email. To most people the idea that a deposed prince of a country halfway across the world wants to give them money in an implausible scenario is obvious fake, however because those emails are sent to so many a few are bound to fall for it. For this reason, I have created PhyFinder, an application that attempts to find phishing emails so that those most vulnerable in our society do not fall for them, costing us millions of dollars every year.

There are many possible characteristics of a phishing email, and as methodologies used by phishing attackers are discovered and prevented, new methodologies are created. In this way unless I have access to a learning model and many thousands of emails as input it is quite near impossible for me to create a perfect and constantly evolving machine. Because of this I attempted to find characteristics used by many phishing emails and create a system that can detect those characteristics in example emails, both created by me and real-life phishing emails. The characteristics I chose to focus on were, spelling errors, suspicious links, evocative language, and the address itself compared to where the email claims it is from. As outlined in the 2006 paper by Chandrasekaran and et. To do this I created a class that sits in front of an inbox that analyzes each email as it comes in, giving it a score from 0 to 100 based on how likely it feels that that email is phishing for information from a user. If an email receives a high enough score, then that email is sorted into the spam folder and the user is notified of the possible attack. In the class sits a function called checkMail that assigns a score to the email on the previously listed characteristics, as well as other information such as the address against known malicious and benign actors. It then takes the highest 75% of those scores and averages them to find the final score. It also keeps track of all emails that receive a high enough score to be considered suspicious, and if it receives enough emails from that source, will filter all future emails from that source directly into spam. The goal of only taking the top 75% of the scores is that an email may contain some of these characteristics but not others. An email claiming to be from your bank may not have many common misspellings but may use urgent language and suspicious link, while an email claiming to be from a foreign prince may have many misspellings but no link or urgent language.

While not perfect I feel like PhyFinder is a great example of how to look at possible phishing messages, looking at a variety of different types of variables and real world attacks used today.

Sources:

https://www.consumer.ftc.gov/articles/how-recognize-and-avoid-phishing-scams

https://www.cynation.com/the-most-effective-phishing-phrases/

https://www.telegraph.co.uk/technology/microsoft/9346371/Nigerian-scam-emails-deliberately-implausible.html

https://7b0c2ec3-a-62cb3a1a-s-sites.googlegroups.com/site/madhuchandr23/home/madhuc-structural.pdf?attachauth=ANoY7crR0wH3Vk0-KtzeK4dZSTkqE9R9umfssC331E84meTkF24xWhPLWghdNqBloJWnXVWmHlhaP-si0noME5ewCke9PWY_p11H5KVYvu9F7VJfyxf6sPRJ0QwX338RlICuprQbsBk9Ovxtj9XwuFf4x12lh7rNL2dvOpKyxckKsTTMr3SVO5801mFckNWsvgPz09z5HBdjTTx6VrAtHB7e3hgIPD0E640P6ax8HCzk8ztGpISTnY%3D&attredirects=0