Introduction

One of the most dangerous threats to the information system at all is social engineering. The emotional manipulation for the target. The weakest part of an information system often is the human factor. In most cases for the social engineer hacking the emotional state of people is much easier that the hardened computer systems. Any type of data that the social engineer can collect from a normal conversation with a human victim can be relevant for him to gain access to his target like guessing some personal password for example. The use of modern technological devices has been of a great benefit for the attacker to gain more access and reach for his target.

Social engineering attacks depends on a some type of communication between the attacker and a worker for example in order for the attack to collect sensitive data filtered from the conversation, that data can be of a great help for the attacker to guess a password, a credit-card number ,or know a name of a coworker which can identify that someone is a potential target for the attacker to operate on with chain attacks and in both cases this also might support larger attacks. For the attacker to get his target’s data, he would convince the victim to do some action like going to a suspicious website or a fake website to enter his confidential details. Other types of that would require the attacker to impersonate a character that the user can trust and pretend it is him on an SMS, Email, or even a phone call, and the victim will be easily manipulated. Surprisingly this is effective and encourage the attacker to use it more frequently.

One of the most popular attacks is phishing attacks, which ask for a reaction from the victim. The attacker is being disguised as popular website banks, administrators from IT departments. It starts sending emails in perfect format to the victim. These emails lead the user to interact by clicking and then this device starts to download malware or open a form that he must fill in with his personal information.

In previous research related to phishing email detection, such as the phishing website filters built into Microsoft Internet Explorer and Mozilla Firefox. Another related previous work was training individuals on social engineering attacks in order to make them more aware and resistant in the future, and this is considered more effective as individuals are always the weakest layer in the system, but it costs more time and effort.

There is another type of attack that is no less dangerous than phishing attacks: it's not depending on emails but includes calls and texting, and in-person communication. Email communication is not real-time, Email communication is not real-time, the attacker doesn't expect he will receive the response immediately, or may the attacker never respond. But the non-email way being in a real-time involved two-way conversation.

The conversation is better for the attacker because it pressures the target to respond without spending time considering the consequences, and that's made from the known-email attacks more efficient.

Existing methods for automatically detecting social engineering attempts are primarily focused on phishing emails.

These methods rely primarily on non-content metadata present in emails, such as contained hyperlinks and SMTP headers, for analysis. Non-email social engineering assaults that aren't associated with accessible information aren't detected by these techniques. When you don't have any metadata to rely on.

There are content-based techniques that look at things like character frequency and word/n-gram frequency in the material. These methods, on the other hand, do not use semantic analysis to extract the meaning of the text and the attacker's aim. The application of existing content-based metrics alone, without semantic analysis, would result in low precision and accuracy.

Related work

The weakest point of a technological system is the people. The psychological manipulation of people in order to gain access to a system for which the attacker is not authorized is often easier than hacking into the hardened software system. And this is the problem we need to solve, a commonly used approach to solve this issue is to make them aware. Training-based training pushes employees to be aware of the consequences of each step they make in their communication. This type of training approaches relies on the human to prevent attacks manually. But this approach still has its vulnerabilities because the human has his vulnerability to social engineering, and it depends on the human’s emotional state of the time of attack.

People who receive that type of social engineering training is expected to have more self-disciplinary awareness by following specific steps to answer any answer from external sources, like “Is this answer confidential to our organization?”, “Should that person ask for this information? “, “Who should have the privilege to ask this question?”, by following these steps the employee should be able to control the communication with external agents. Maintaining to keep following these questions is often hard especially when the employee is under the emotional influence of the social engineer.

For automatic detection of suspicious emails there is a number o approaches to detect phishing emails using email headers, SMTP headers NIDS logs and others. What we are interested in is the approaches that are focused on processing email content or parts of it. One way of doing that is preforming authorship of identification, to analyze email features and assuring that who claimed to be the sender of the email in the content is really the sender. These approaches characterize emails by using statistics on frequency of words using n-grams. One example of that is ASCAI which generates a writeprint of senders, then generate another writeprint for the current email sender, and compare against the list of known senders’ writeprints to verify authorship.