A Review of Factors Affecting the Effectiveness of Phishing

Robert Karamagi[1]

[1] Department of Information Communication Technology, The Open University of Tanzania, Dar es salaam, Tanzania

Correspondence: Robert Karamagi, Department of Information Communication Technology, The Open University of Tanzania, Dar es salaam, Tanzania.

Received: October 29, 2021 Accepted: November 16, 2021 Online Published: November 30, 2021

doi:10.5539/cis.v15n1p20 URL: https://doi.org/10.5539/cis.v15n1p20

Abstract

Phishing has become the most convenient technique that hackers use nowadays to gain access to protected systems. This is because cybersecurity has evolved and low-cost systems with the least security investments will need quite advanced and sophisticated mechanisms to be able to penetrate technically. Systems currently are equipped with at least some level of security, imposed by security firms with a very high level of expertise in managing the common and well-known attacks. This decreases the possible technical attack surface. Nation-states or advanced persistent threats (APTs), organized crime, and black hats possess the finance and skills to penetrate many different systems. However, they are always in need of the most available computing resources, such as central processing unit (CPU) and random-access memory (RAM), so they normally hack and hook computers into a botnet. This may allow them to perform dangerous distributed denial of service (DDoS) attacks and perform brute force cracking algorithms, which are highly CPU intensive. They may also use the zombie or drone systems they have hacked to hide their location on the net and gain anonymity by bouncing off around them many times a minute. Phishing allows them to gain their stretch of compromised systems to increase their power. For a normal hacker without the money to invest in sophisticated techniques, exploiting the human factor, which is the weakest link to security, comes in handy. The possibility of successfully manipulating the human into releasing the security that they set up makes the life of the hacker very easy, because they do not have to try to break into the system with force, rather the owner will just open the door for them. The objective of the research is to review factors that enhance phishing and improve the probability of its success. We have discovered that hackers rely on triggering the emotional effects of their victims through their phishing attacks. We have applied the use of artificial intelligence to be able to detect the emotion associated with a phrase or sentence. Our model had a good accuracy which could be improved with the use of a larger dataset with more emotional sentiments for various phrases and sentences. Our technique may be used to check for emotional manipulation in suspicious emails to improve the confidence interval of suspected phishing emails.

Keywords: phishing, social engineering, machine learning, emotion detection

1. Introduction

Phishing is a form of social engineering, and an unethical act performed by malicious adversaries that aim to capture sensitive user information through manipulating human emotion. Attackers have so many tricks up their sleeves that they use to get victims to give up their credentials, or bank information online. They normally send fake links to them in the hope that they may be intimidated to see where they may take them. These links direct the victims to evil websites, which leverage on the users' anticipation, sympathy, fear, or any other human-like factor to steal their confidential information, which they may later use to gain unauthorized entry into their systems.

Once inside their systems, the hackers may gather the information that is even more critical. This information may allow the hackers to always hack the victim's machine (persistence), or be able to break into their other devices, social media accounts and applications. Most of the time, attackers use emails, instant messaging, etc. to get their evil links to the victims. Because new techniques are always being devised by hackers, social engineering awareness training by far serves as the best means to prevent phishing attacks. (Saxena et al., 2019)

Social engineering

Social engineering is a non-technical type of attack, but it can be used in collaboration with technical attacks such as spyware, keyloggers, backdoors, remote access trojans (RATs), viruses, rootkits, etc. Not much effort is needed to convince humans to give away personal or corporate information, which may be very useful to an intruder. Social engineers do not necessarily use sophisticated or normal technical attacks, but they achieve their goals by psychological intelligence and crafty social opportunism.

Nowadays, business firms are investing a lot of money in cybersecurity solutions. Never-the-less, a naïve employee may compromise these expensive solutions and give hackers a way in without even knowing, and with no trouble or efforts to bypass the security intelligence from the hackers. This is the essence of social engineering. It is about targeting the weakest element in an organization's security, which is the human factor, as they can be hacked much easier in comparison to computers. Ethical people are normally kind and help others in need, which is what makes them so vulnerable to the unethical ones. (Abass, 2018)

Spear Phishing

Spear phishing is a form of phishing attack that targets a single primary individual. The attacker focuses their attention on a single person and lures them into surrendering their secret data without them having a clue what is going on. The email consists of information concerning the target directly such as their home address, the company they work for, their job title etc. Social media sites such as LinkedIn have become a useful tool for hackers to plot their spear phishing hacks due to the professionalism aimed by users in submitting true information relating to their academic history and professional careers. (FireEye, 2018) (Bullee et al., 2017)

Vishing

Voice phishing or Vishing is another form of phishing where by the attacker makes a phone call to the victim and then tries to manipulate them over the phone in giving up confidential data. Hackers find vishing convenient as they may avoid detection by bouncing off several cellphone towers making them hard to trace. Reports have revealed that most victims of such attacks do not file reports to the police or responsible authorities. (Maseno, 2017)

Whaling

Whaling is a form of phishing that targets the high-profile individuals such as chief executive officers (CEOs), presidents, kings or queens etc. The attackers perform intensive profiling for extended periods before initiating an attack. This is a relatively serious type of threat as top executives have access to the most sensitive and critical data. A loss of such information is taken to have grave consequences. (Gupta et al., 2018)

Keylogger

A keylogger is a form of spyware that can be either hardware or software based. Its sole purpose is to record all the keystrokes that are typed in by the user, without them knowing. It does this behind the scenes so the user cannot suspect anything. The strokes of each key are compiled altogether into a log type of file. The hacker makes it so the logs may be sent to them secretly via email or any other covert mechanism. (Parekh et al., 2020)

Backdoor

A backdoor is a form of malware that functions to allow hackers to gain access to a machine without any approval, authorization, or authentication. The attackers normally use them to circumvent authentication systems and gain unapproved remote-control sessions to a machine or device. Once in a machine, they can hide themselves in many ways and go on further to steal confidential data or cause a denial of service. (Loi, 2017)

Remote Access Trojan

A remote access trojan is a type of backdoor that allows a hacker to remotely take full and unabated control over a device. It is installed without the users' awareness or consent and hides to avoid being detected. The hacker is capable of running a large number of different evil commands and can take control of the webcam, microphone, operating system, peripherals and many more. It may behave as spyware to capture keystrokes, audio, and video in the background and ship it to the hacker in real-time. (Valeros & García, 2020)

Reverse Shell

A reverse shell is a connection shell that is virtual and open to the attacker's machine that initiates from the victims' device. With an open connection, an attacker is capable of running scripts and executing various evil commands on the victim's machine. This mechanism is normally used by many common RATs and backdoors. The hacker gains the privileges of the user who was logged in when the session was initiated and may work to elevate privileges to gain further access. Most reverse shells use the Transmission Control Protocol (TCP) for the establishment, but the Internet Control Message Protocol (ICMP) has also been observed in use. Any port may be used to create the communication link. When hiding under allowed ports, the firewall and other intrusion detection systems face difficulty in identifying an incident because the ports are approved to communicate. When the shell is run under port 443 Secure Sockets Layer (SSL), content inspection is very hard as the traffic is encrypted. (Lu, 2019)

Virus

A computer virus is a dangerous piece of code that self-replicates by infecting other programs within its reach by injecting malicious code in them. Many viruses are dependent on some kind of executable, of which they are hosted by, before they may be able to start running their code. They may become very severe, to the point where they may completely ruin all of your hardware and software. A virus cannot propagate without human intervention. This is what makes it different from a worm. A virus needs some kind of host program to keep it in action. Viruses may be spread very rapidly and unintentionally as victims may tend to share poisoned files or send attachments in the email that have been infected. (Kumar & Dey, 2019)

Worm

A computer worm is a malicious program that replicates on its own without the need of interacting with any other file, and spreads across machines in a network. It transmits copies of its code throughout the network without attaching to any programs like in the case of a virus. They normally consume a large amount of the network bandwidth and cause a lack of service availability to users on the network. The first case of a worm propagating over the internet was the Morris worm released by Robert Tappan Morris on November 2, 1988. (Jajoo, 2017)

Rootkit

A rootkit is a program written primarily for evading detection while maintaining privileged access to the system. Host Intrusion Detection Systems (HIDS), Host Intrusion Prevention Systems (HIPS), Network Intrusion Detection Systems (NIDS), Network Intrusion Prevention Systems (NIPS), firewalls, Security Information and Event Management (SIEM) solutions all find it very hard to detect rootkits as they can modify their underlying operating system kernel code that they utilized for detection and thus usurping the control of security. (Nadim et al., 2021)

Despite the various technical attacks which a hacker may deploy on a target, psychological attacks are the most dangerous. Victims' emotions are the most vulnerable asset in a phishing operation and it is the attacker's main priority. To efficiently be able to reduce the effects of such an attack, it is essential to ask ourselves, what emotion are they trying to break? Our study focuses on applying machine learning to train a model to be able to classify the emotion present from a group of words. The model is trained using a large dataset of social sentences and phrases with differing emotions. We match positive emotions to positive sentiments, and vice versa. Likewise, neutral emotions are matched to neutral sentiments. The model is trained against the resulting dataset of emotions and related text.

Section 2 is a literature review section where we shall look at various contributions to detect, improve awareness, or counteract phishing. Section 3 is the methodology section. Insights of the attacking methods hackers use to achieve their phishing attempts are discussed here. The implementation methodology of our emotion detection scheme to capture the emotion attackers exploit in their phishing email is explained. Section 4 is the results and discussions section. We talk of the findings of what factors motivate hackers to perform phishing and make victims more susceptible to phishing attacks. We also look into how our emotion detection scheme performs. Section 5 is the conclusion section where we conclude our exploratory research review and propose future studies to enhance phishing detection schemes.

2. Literature Review

Jampen et al. conducted a survey to analyze the effectiveness and sustainability of organizational training programs, aimed at giving anti-phishing awareness to its employees, in reducing their employees' vulnerability to the phishing attacks. They categorized works using a well-formed methodology that took into consideration various parts of the training programs. They noted important results in the technical literature. They found out that, overall, researchers found a consensus to most research questions that gave regard to the convenience of training programs for anti-phishing. A mixture of findings came from how age affects the likelihood of the accomplishment of anti-phishing programs to train employees. Jampen et al. gave a description based on their comprehensive analysis on the design of a properly structured training program for anti-phishing and a framework with a set of recommended directives for research. (Jampen et al., 2020) Their study reveals how phishing awareness training can reduce the attacker's chances of success, which will need them to go the extra miles in convincing users who have been trained already.

Panum et al. performed an exploratory work on the efficiency of modern and popular and solutions that detect phishing, by analyzing the techniques of detection that they shared in common. They presented sample mechanisms capable of avoiding detection by causing unnoticeable perturbations. They proposed steps and measures to take in the design to improve the evaluation of the robustness of adversaries in the future. They brought to light a terminology, for respective methods, that does not depend on the application or environment to elucidate the conditions for the setting of the adversaries. Three axioms that should be accounted by any solution that detects phishing attacks were presented by them, based on an agreed upon phishing definition. This aided the solutions not to use the wrong inference attributes. They disintegrated the inference methods from the detection solutions into a collection of strategies. They evaluated the capability of absconding capture and cases of perturbations that allowed it. Their findings allow the definition of guidelines to the design of solutions for phishing detection for the community. (Panum et al., 2020) This means that hackers are sure to face impedance as they design common phishing websites, as phishing solutions equipped with such guidelines shall be detected.

Bitaab et al. carried out an extended study of measurements on the attacks related to phishing that occurred during the premature stages of the coronavirus pandemic between January 2020 and May 2020. They used their dataset to perform tracking of the trends and the reason for the growth in the phishing attacks. They analyzed and collected records of Domain Name Systems (DNS), certificates of Transport Layer Security (TLS), the phishing websites' source codes, web traffic, and Uniform Resource Locators (URLs), emails of phishing attacks, news, and announcements of the government. They found out that the traffic of phishing attacks increased by a fraction of 220% in comparison to the rate before the COVID-19 pandemic, which exceeded previous seasonal trends. The attackers would orchestrate various hacks to manipulate the victims' uncertainty and fear about the pandemic. The attackers used modern ways to which the existing defense systems could not handle. The analysis of Bitaab et al. displayed the ability for new defenses of ecosystems and upgraded teamwork among parties to promote quicker and efficient plans for ecosystems to tackle the uprising volume of phishing. (Bitaab et al., 2020) Their study provides quicker mutations to phishing defenses, which may be able to tackle unplanned scenarios such as a pandemic. This may reduce the hacker's ability to take advantage of uprising tragedies to perform their cybercrime.

Steves et al. proposed a rating scale to solve the issue faced by organizations investing in training programs for phishing awareness. Since Chief Information Security Officers (CISOs) are highly dissatisfied if phishing click rates that result from the training exercises are very high. The training budget must be justified to the board officials explaining the necessity of the training as the click rates are not reducing. Their study gave rise to a debate that the level of difficulty of the phishing attack email targeted at an individual should be a factor in measuring the variance of the click rates. Based on previous studies, a phishing email forged to align with the context of a user's job is much harder for the users to detect malice. This made Steves et al. come up with a Phish Scale to aid CISOs and the implementers of the phishing tests to give a rating of how difficult their attack is. The scale justifies any associated clicks from the tests. Cues of phishing and the context of users from former researches devised the base of their scale. They applied the scale to recent and old published results from the phishing attack exercises performed by enterprises. Their Phish scale had decent results with their selected datasets and revealed large potential as a scaling tool and a catalyst for sharing information regarding the clicks observed during a phishing experiment. (Steves et al., 2020) Their tool provides disambiguation to the top management as phishing training clicks rise in number, which helps them be able to make more effective decisions. This ensures that phishing training may have a stable budget and hackers shall more likely have to face highly trained staff, which again may reduce their chances for success.

Wang et al. attempted to detect phishing using Bidirectional Long Short-Term Memory (BLSTM) and Random Forest classifiers. The results of their experiments were satisfying in regards to the detection of phishing and their study contributed to applying the algorithms that they proposed to the field of information security. The Bidirectional Long Short-Term Memory model produced a rate of recognition of 95.47% in comparison to the Random Forest model that produced 87.53%. The results of Wang et al. reveal that the BLSTM detection method is more reliable in guaranteeing security in the network and uncovering the relevance of the model that they proposed to detect phishing. (Wang et al., 2020) Their study revealed efficient means to tackle phishing, and future solutions may be adapted to fill gaps in weaknesses of solutions of the past.

Mohith Gowda et al. devised a novel mechanism to identify websites used for phishing with the use of a novel architecture for a browser on the client's side. They applied the extraction framework rule to draw out the websites' properties using only the URL. A list consisting of 30 various features of a URL is populated. The

Random Forest Machine Learning Classification Model uses it later on in detecting if the website is authentic. 11,055 tuples were in the dataset used for the model training. Courtesy of the reconfigured architecture of the browser, the processes were capable of allowing the client's side to perform them. Because most normal users do not have proficient technical knowledge on the usage of manual frameworks for phishing site detection based on machine learning, they improvised to make sure any individual accesses their tool. Mohith Gowda et al. developed the 'Embedded Phishing Detection Browser' (EPDB), serving as the method for detecting the phishing sites embedded within the architecture of the browser, to maximize user experience and enhance security simultaneously. Their novel architecture runs the tasks for phishing detection dynamically in real-time and was accurate by 99.36% in identifying the malicious phishing sites. (Mohith Gowda et al., 2020)

Aljofey et al. demonstrated a phishing detection solution that utilizes a character-based Convolutional Neural Network (CNN) to examine the URL of the website using a model based on fast deep learning solutions. Their model incorporates neither using any services from third parties nor retrieving content from the target website. They captured sequential and information patterns of the URL strings without the need of an idea about phishing beforehand. The sequential pattern features fast classify the original URL. They also compared various traditional and deep machine-learning models. Features sets included crafts by hand, embedded characters, Term's Frequency-Inverse Document Frequency (TF-IDF) and count vectors features at the character level. The experimental results of Aljofey et al. brought an accuracy of 95.02% on their dataset from the model that they proposed. Benchmark datasets that perform better than the present models of phishing URLs produced accuracy scores of 98.58%, 95.46%, and 95.22%. (Aljofey et al., 2020)

Sharma & Bashir performed a study where they looked into the attackers' emails that were available to the public in repository databases for phishing attacks. They analyzed the characteristics and contents of those phishing emails. In order for them to understand the language and techniques that attackers used to be able to lure their targets, they considered many variables. Their findings showed that the words of the attackers used in their emails would aim at exploiting emotional triggers in humans such as anticipation and fear. The role played by their findings centered on a human study is a major step directed to improving the programs for training and enhancing the detection of phishing attacks, similarly, human factors may take part in the security of systems. (Sharma & Bashir, 2020)

Broadhurst et al. performed a study based on explorations and observations on 138 recruits from the orientation week of a university for several months in 2017. The aim of their study was to find out cybercrime risks. They ran social engineering attacks, observed the responses, and compared how the participants took the risks to cybercrime before and after the phishing campaign. Their quasi-experimental survey exposed the test subjects to fake emails and phishing attacks. The intention of the emails was to steal confidential data from the victims or convince them to navigate to poisoned sites by clicking on the malicious links in the mail. Their techniques varied in terms of individualization. The phishing categories involved targeted or spear, tailored and generic. They classified the subjects based on the awareness of cybercrime in two groups, viz. Hunter and Passive condition. Those in the Hunter class, throughout the experiment, were aware of all forms of swindles and the ones in the Passive class did not get any warning. Broadhurst et al. analyzed the effects of the type of scam, awareness of cybercrime, competence in the field of information technology (IT), gender of the subject, and perception of safety on the internet to how susceptible their email scams were. They found out that spear phishing had a better chance of being engaged to than a generic phishing attempt. Their analysis also pointed out that there was a higher probability that fresh men and international students would face deception from the phishing than the senior and domestic students. For the further exploration of all their variables and the results, they performed a generalized linear model (GLM) analysis. (Broadhurst et al., 2020)

Williams & Joinson conducted theoretical based research to investigate the methods in which the present and future phishing interactions may target the users along with the effect they play on the susceptibility of phishing. They developed and validated a survey measure centered on the protection motivation theory constructs across two studies. Such constructs include perceived vulnerability, efficacy to response, severity, and self-efficacy. They assessed the features contributing to the decision that people make on whether they shall stay updated by phishing awareness to protect themselves or not. They analyzed what role each construct played in the intentions of the user to know the latest phishing techniques that will evolve and the capability of phishing discrimination via a phishing quiz assessment. Williams & Joinson observed that larger intentions came from a greater perception of the threats' response efficacy, severity, and self-efficacy while low intentions to know about the phishing techniques came from a high perception of vulnerability. They did not manage to find a relationship with the ability to discriminate against phishing. With the knowledge on the causes of users' intentions in maintaining education and pursuing updates about phishing dangers, the assurance is available, that efficient interventions come, and maximum potential effects exist. (Williams & Joinson, 2020)

Frauenstein & Flowerday presented a model based on theory to counter the susceptibility of phishing on social network sites (SSNs). They collected data from 215 subjects and observed the contribution of the processing of information to phishing on social networks. They regarded how users are vulnerable to the sites based on their personalities to identify characteristics of users that may be more susceptible to phishing on these social media sites. They performed a Structural Equation Modeling (SEM) analysis, and the results showed that heuristic processing faced a negative impact from the conscious users, and thus were less vulnerable to phishing on SNSs. Their analysis supported and confirmed previous studies that the susceptibility to phishing increases with heuristic processing. The research of Frauenstein & Flowerday contributed to the discipline of information security as being one of the first studies to analyze how the relationship between the Big Five Personality Traits: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness, and the systematic heuristic model is affected. (Frauenstein & Flowerday, 2020)

3. Methodology

The Cyber Kill Chain® framework, developed by Lockheed Martin, is a constituent of the Intelligence Driven Defense® model. This model is used to identify and prevent cyber invasion activities. It tells us what threat actors must do to become successful in their attempt to break into a system or network. Visibility into the attack is increased and a security, forensic, or risk analyst is capable of understanding the techniques used by an attacker to gain forbidden or unauthorized access. (Spitzner, 2019)

1. Reconnaissance

This is the stage where we go through the social media profiles of our victim that are available publicly on the Internet such as Twitter, Instagram, Facebook, LinkedIn, Snapchat, Telegram, etc. We look for any useful information that may aid our attack such as friends, family, and pet names, hobbies, education, profession, work experience, and the like.

2. Weaponization

At this point, we may create an evil payload such as an application that is embedded with a backdoor virus that we aim to ship off to our victim and get them to run.

3. Delivery

We use the phishing social engineering technique to pretend we are someone with the ability to convince the victim into doing our malicious deed. We trick the victim into fetching our weaponized payload and storing it on their system.

4. Exploitation

The technique that made the virus reach the victim’s machine; whether it was manipulating the trust, fear, sympathy, or any other human emotion of the victims through phishing, is the root cause of the virus eventually being executed and the exploit becoming successful.

5. Installation

Once the exploit is successful, spyware, malware, and other dangerous programs can be installed to give an upper hand in gaining elevated privileges and having a persistent route to breaking into the victim’s machine at any time in the future.

6. Command & Control

Commands may be run to capture secret data, and learn critical information to achieve greater hacks. The system may be controlled remotely and the firm power of the victim’s machine is gained.

7. Actions on Objectives

At this stage, it is all up to the limit of the malicious intentions devised at the start of the chain that will determine the course of actions and further proceedings.

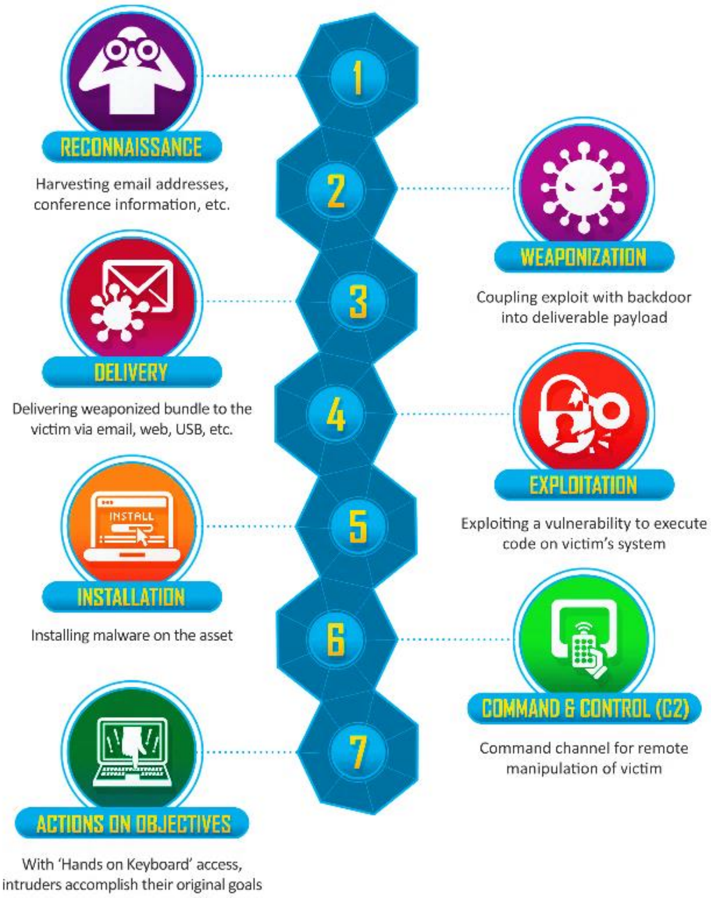

1
RECONNAISSANCE
Harvesting email addresses, conference information, etc.
2
WEAPONIZATION
Coupling exploit with backdoor into deliverable payload
3
DELIVERY
Delivering weaponized bundle to the victim via email, web, USB, etc.
4
EXPLOITATION
Exploiting a vulnerability to execute code on victim's system
5
INSTALL
INSTALLATION
Installing malware on the asset
6
COMMAND & CONTROL (C2)
Command channel for remote manipulation of victim
7
ACTIONS ON OBJECTIVES
With 'Hands on Keyboard' access, intruders accomplish their original goals

Figure 1. The Cyber Kill Chain® framework (Lockheed Martin)

3.1 Email Phishing

We may classify the techniques of phishing attacks into two basic classes. The first is attack launching and the second is collection of data. There are various techniques used in attack launching such as spoofing emails, malicious attachments, social settings abuse, spoofing URLs, spoofing websites, voice phishing, working in collusion in a social network, man in the middle (MiTM) attacks, spear phishing etc. Data collection methods involve post intrusion techniques where all important and useful information is gathered after gaining entry. This data may be used to launch further and more lethal attacks. The data collection can be automated or manual. Automatic techniques involve fake websites and keyloggers while manual techniques are those like misdirecting a human and social masquerading. Figure 2 below depicts a phishing attack diagram. (Basit et al., 2021)

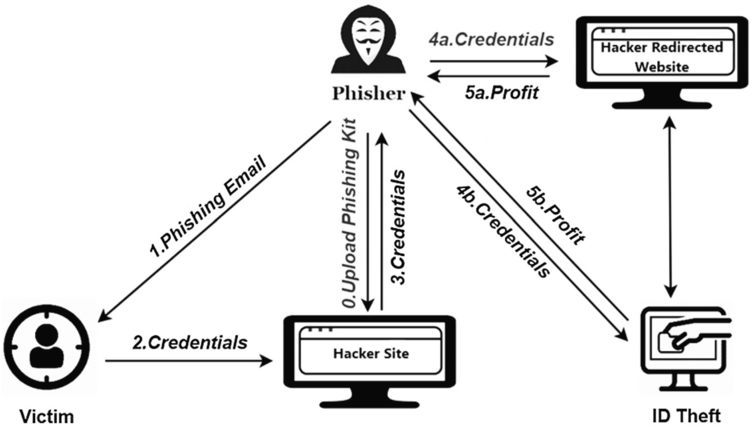


Figure 2. Phishing attack diagram (Basit et al., 2021)

A phishing attack based on email is shown in Figure 3 below. Attackers start by observing various organization sites across the internet. They find out a specific website that requests all the data that they would desire to steal from their victims. Once they have the site in mind, they clone it using autonomous scrapers and crawlers. A very common website copier is HTTrack. The hackers send an email to victims encouraging them to go and input their details in their malicious fake site. If the email sounds convincing enough to the victim and they follow the link, and enter their real confidential data, thinking it is a legitimate website, they would have simply handed over their secret information right into the hands of the hackers. (Jampen et al., 2020)

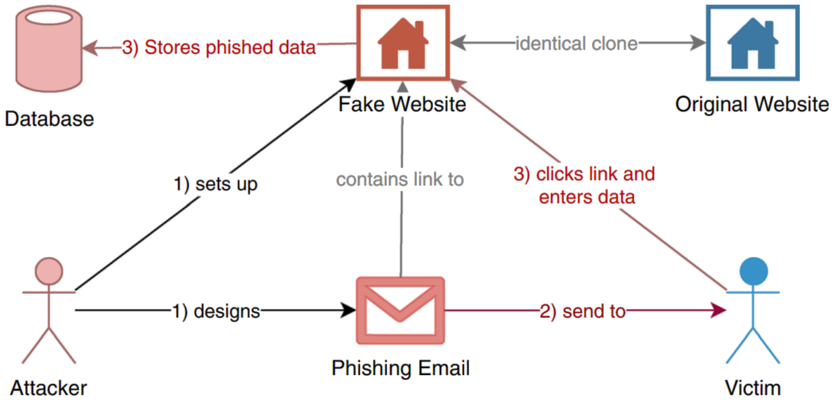
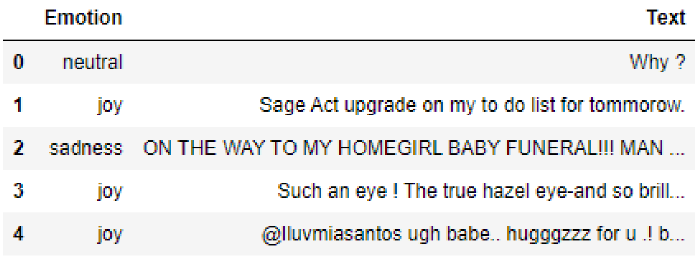


Figure 3. Email based phishing attack (Jampen et al., 2020)

3.2 Experimental Work

We designed a machine learning model capable of detecting the emotion of a phrase or sentence which may be used to find the emotion a hacker is trying to manipulate from their phishing email. Our dataset comprised of 34792 unique values with 8 different types of emotions namely joy, sadness, fear, anger, surprise, neutral, disgust, and shame. Figure 4 shows the head or first five rows of our dataset.



	Emotion	Text
0	neutral	Why ?
1	joy	Sage Act upgrade on my to do list for tommorow.
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...
3	joy	Such an eye ! The true hazel eye-and so brill...
4	joy	@lluvmiasantos ugh babe.. hugggzzz for u .! b...

Figure 4. First five rows of emotion dataset

The bar plot of the emotions against the total count of our dataset is illustrated in Figure 5. We performed a sentiment analysis to find out the sentiment associated with the text and emotion. 3 sentiments were matched with the data, namely, positive, negative, and neutral. Figure 6 shows the first five rows of our dataset with sentiments of the text included.

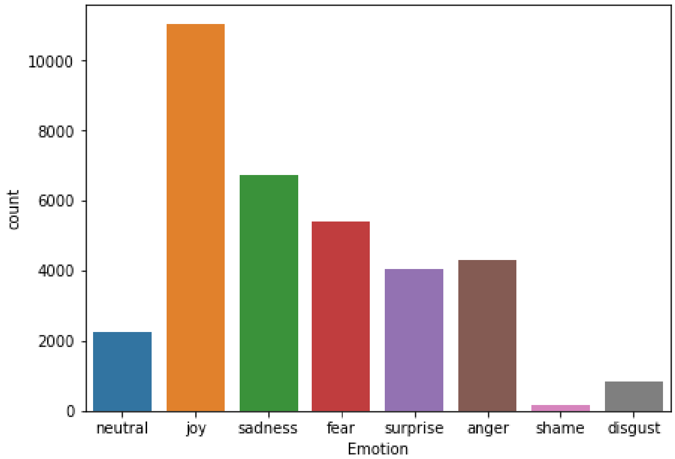
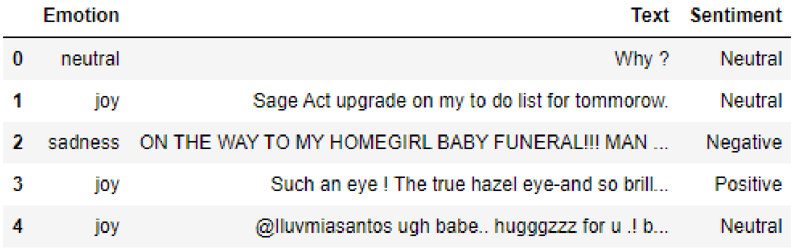


Figure 5. Bar plot of emotions count



	Emotion	Text	Sentiment
0	neutral	Why ?	Neutral
1	joy	Sage Act upgrade on my to do list for tommorow.	Neutral
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...	Negative
3	joy	Such an eye ! The true hazel eye-and so brill...	Positive
4	joy	@Iluvmiasantos ugh babe.. hugggzzz for u .! b...	Neutral

Figure 6. First five rows of dataset with sentiment column

We enriched the quality of our dataset by filtering the data to have the same type of emotion and sentiment. We matched positive emotions with positive sentiments and vice versa. For example, we considered only positive sentiments from the joy and surprise emotions, and negative sentiments from the sadness, fear, anger, shame, and disgust emotion. Similarly, the neutral sentiments only were taken from the neutral emotion. Figure 7 shows the distribution of each emotion and their respective sentiments. Figure 8 shows a sample word cloud of the joy emotion from the keyword extraction process, where we extracted the most common words per class of emotion.

We performed text cleaning on the resulting dataset to remove noise variables such as stop words, special characters, punctuations, and emojis. We used 70% of the cleaned data for training our model using logistic regression across 1000 maximum iterations.

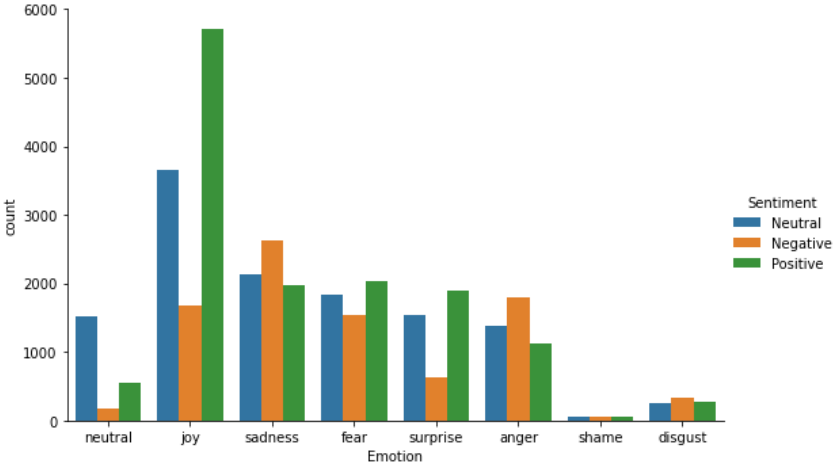


Figure 7. Distribution of emotions and sentiments

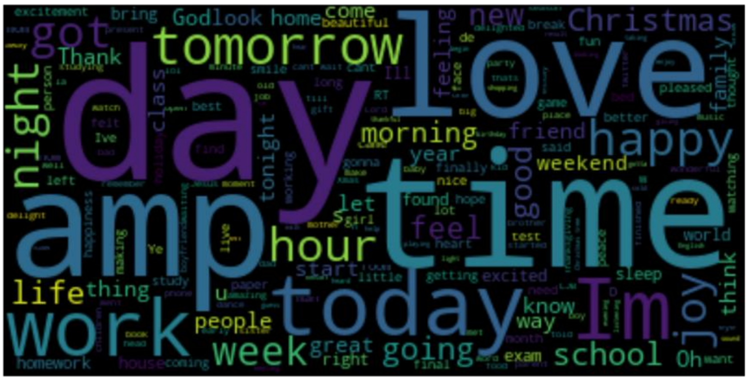


Figure 8. Word Cloud of joy keywords

4. Results and Discussions

4.1 Factors Affecting the Motivation of Hackers to Perform Phishing

The success of phishing attacks is primarily due to the lack of awareness about it in society. Hackers are mainly motivated by financial gains as they invest time, money and thoughts in performing their hacks. Social gain may also be their goal in performing such crimes. (Gupta et al., 2018) Such scenarios involve-

- Stealing the credit card data such as card number, card verification value (CVV), expiry month/year.
- Personal information that is stolen may be sold online to the highest bidder. Data such as telephone numbers, biography, demographics etc. are in high demand in the darknet.
- Spear phishing may be applied to steal classified documents and trade secrets from targeted companies and sell them off to competitors or any other buyers.
- Sometimes hackers are not financially motivated but do it for fame and social status to get respected.
- The bad people are usually also interested in finding out where vulnerabilities exist in systems so they can use them to launch exploitation attacks later on in the future.
- Hackers also need computing resources such as central processing unit (CPU) and random-access memory (RAM) so they would like to hack and hook computers into a botnet. This may allow them to perform distributed denial of service (DDoS) attacks and perform brute force cracking algorithms, which are CPU intensive. They may also use the zombie or drone systems they have hacked to hide their location on the net and gain anonymity by bouncing off around them many times a minute.

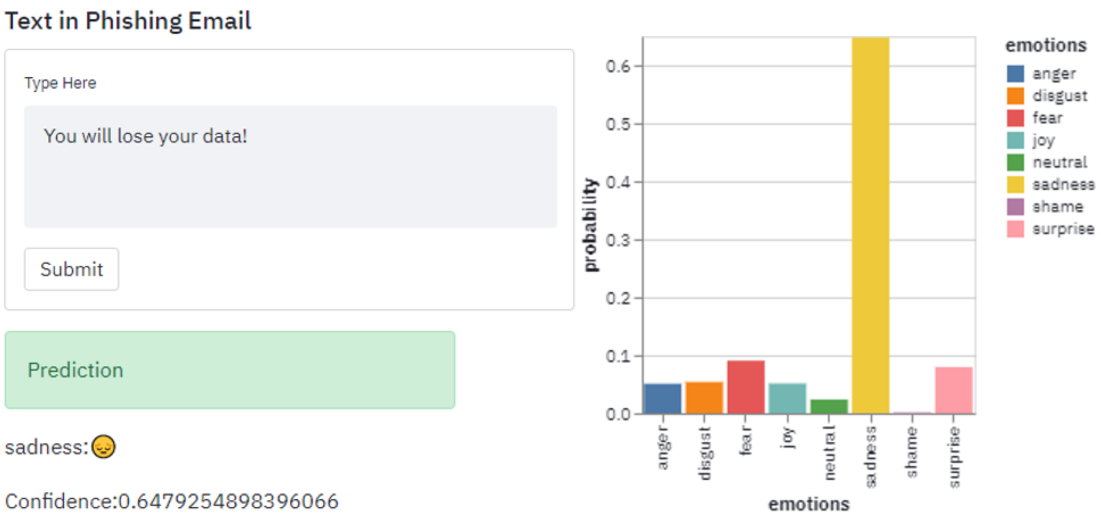
4.2 Factors Affecting the Susceptibility of Victims to Phishing

- *Curiosity.* The hackers try to convince the victims by talking about breaking news or interesting facts that the victims may like, which they found out from the social networks.
- *Fear.* Emails telling the users that there may be some negative consequences to their possession or loved ones entice fear and cause them to navigate to the malicious pages.
- *Empathy.* The attackers may pretend to be a friend or family member that needs help to convince the victim to visit their infected website. (Broadhurst et al., 2020)
- *Authority.* It is normal for a human to conform when a high authority is giving instructions to them.
- Commitment. The human character of pursuing an objective till the end once they have started it can allow hackers to indulge them to follow their directions.
- *Liking.* People may be easily convinced to do something by someone they care about or like. The hacker may find out this person from social media and pretend to be them.
- *Contrast.* An attacker may form a manipulative email, which has two contrasting options, so in the event the victims disagree with one option, they may be encouraged to commit to the other opposite one. However, all options have been maliciously contaminated.

- *Reciprocity.* When living in a society, people enjoy returning any good or kind acts, which have been done to them. An attacker may pretend they are asking for the return of a favor that they have managed to find out or create through social engineering.
- *Scarcity.* A hacker demonstrating a lack of availability or uniqueness may trick an individual into perceiving value in their fake email or message.
- *Social proof.* Usually, people feel better doing something if everyone is also doing it. Hackers may use this to their advantage by faking that their issue is a common practice.

4.3 Emotion Detection

We used a Streamlit application program interface (API) to connect our emotion detection machine learning model to a frontend web application. We tested several sentences to get the emotion associated. The predictions could be improved by training our model with a larger dataset having more instances of text for the various emotions. Figure 9 below shows the sadness emotion detected with a 64.7% confidence, from a sentence, “You will lose your data!”



Text in Phishing Email

Type Here

You will lose your data!

Submit

Prediction

sadness:😞

Confidence:0.6479254898396066

Figure 9. Emotion detection application

5. Conclusion

In our paper, we have looked at phishing in general and various concepts related to it. We have performed a high-level analysis of the techniques hackers use to penetrate systems with the use of email phishing. We have reviewed various literature from different authors about the ways to prevent phishing using machine learning and neural networks along with various approaches to determine what influences the success rate of phishing. We have seen that the human factor plays a major role in the effectiveness of phishing and security training and awareness is a key technique that may reduce the threat scale and risk of being attacked. We managed to apply the use of machine learning to detect emotions from text which can be used to capture the emotion that a hacker tries to manipulate from the text in a suspected phishing email. Future work involves a qualitative and quantitative approach to evaluate which emotion is more effective in achieving a successful exploit. Qualitative methods may be achieved by a survey questionnaire to get an honest judgement from a possible victim of social engineering. Quantitative methods involve measuring the success rate of phishing exploits across various emotion use cases.

Author

Robert Karamagi is both an Electrical and Electronics and a Computer Science Engineer. He currently works in the Information and Communications Technology Department at CRDB Bank Plc in Dar es salaam, Tanzania. Robert specializes in Information Technology Security and Compliance. His research interests include Computer Science, Cybersecurity, Machine Learning, Penetration Testing, and Robotics.

References

Abass, I. A. M. (2018). Social Engineering Threat and Defense: A Literature Survey. *Journal of Information Security*, *09*(04), 257-264. https://doi.org/10.4236/jis.2018.94018

Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2021). A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommunication Systems*, *76*(1), 139-154. https://doi.org/10.1007/s11235-020-00733-2

Broadhurst, R., Skinner, K., Sifniotis, N., Matamoros-Macias, B., & Ipsen, Y. (2020). Phishing risks in a university student community. *Trends and Issues in Crime and Criminal Justice*, *2*(587), 4-23. https://doi.org/10.52922/ti04251

Bullee, J., Montoya, L., Junger, M., & Hartel, P. (2017). Spear phishing in organisations explained. *Information & Computer Security*, *25*(5). https://doi.org/10.1108/ICS-03-2017-0009

FireEye. (2018). Spear-Phishing Attacks: Why they are successful and how to stop them. *White Paper*.

Gupta, B. B., Nalin, A. G. A., & Psannis, K. (2018). Defending against Phishing Attacks: Taxonomy of Methods, Current Issues and Future Directions. *Telecommunication Systems*. https://doi.org/10.22363/2313-2272-2018-18-1-117-130

Jajoo, A. (2017). *Term Paper on Morris Worm*. *December*, 1-18.

Jampen, D., Gür, G., Sutter, T., & Tellenbach, B. (2020). Don’t click: towards an effective anti-phishing training. A comparative literature review. In *Human-centric Computing and Information Sciences* (Vol. 10, Issue 1). Springer Berlin Heidelberg. https://doi.org/10.1186/s13673-020-00237-7

Kumar, R., & Dey, S. (2019). STUDY OF COMPUTER VIRUS TRANSMISSION. *International Journal of Research and Analytical Reviews*, *6*(1), 542-549.

Loi, H. (2017). Low-cost Detection of Backdoor Malware. *12th International Conference for Internet Technology and Secured Transactions*. https://doi.org/10.23919/ICITST.2017.8356377

Lu, L. (2019). Detect Reverse Shell Attack. *TriagingX*.

Maseno, E. M. (2017). VISHING ATTACK DETECTION MODEL FOR MOBILE USERS. *KCA University*.

Nadim, M., Antonio, S., Lee, W., City, N. Y., Akopian, D., & Antonio, S. (2021). *Characteristic Features of the Kernel - level Rootkit for Learning - based Detection Model Training*. 1-7. https://doi.org/10.2352/ISSN.2470-1173.2021.3.MOBMU-034

Parekh, D. H., Adhvaryu, N., & Dahiya, V. (2020). *Keystroke Logging : Integrating Natural Language Processing Technique to Analyze Log Data*. *3*, 2028-2033. https://doi.org/10.35940/ijitee.C8817.019320

Sharma, T., & Bashir, M. (2020). *An Analysis of Phishing Emails and How the Human Vulnerabilities are Exploited* (Issue October). Springer International Publishing. https://doi.org/10.1007/978-3-030-52581-1

Spitzner, L. (2019). Applying Security Awareness to the Cyber Kill Chain. *SANS Institute*.

Valeros, V., & Garc á, S. (2020). GROWTH AND COMMODITIZATION OF REMOTE ACCESS TROJANS. *Czech Technical University in Prague*. https://doi.org/10.1109/EuroSPW51379.2020.00067

Copyrights