Deception of Phishing: Studying the Techniques of Social Engineering by Analyzing Modern-day Phishing Attacks on Universities.

by

Lauren Elizabeth Walker

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 7, 2016

Keywords: Social engineering, phishing emails

Approved by

Greg Weaver, Chair, Associate Professor of Sociology
Allen Furr, Professor of Sociology
Robert Norton, Professor of Veterinary Microbiology, Public Health and Biosecurity

Abstract

Techniques applying manipulation, persuasion, and influence have been used for centuries to conduct malicious acts of crime against others. Social engineering is the act of using various manipulation techniques to get other individuals to give private or confidential information they would not otherwise divulge. In today's technologically connected society, social engineers have begun to focus and apply their skills onto the cyber realm in hopes of gaining access to unauthorized or private information. A sample of phishing emails attempts on an American university is utilized to analyze patterns of influence and persuasion techniques of social engineers. By understanding the patterns of the phishing emails hitting college campuses, a generic profile can be developed of what campus officials should prepare for both proactively and reactively to secure their institutions

Acknowledgements

Completion of this thesis could not have been possible without the help of devoted professors and peers. Throughout my time as an undergraduate student and a graduate student at Auburn University, Dr. Greg Weaver has always been more than happy to provide me with experience, advice, and classroom knowledge that has allowed me to grow my specific career interest. A special thanks to Dr. Allen Furr and Dr. Robert Norton for serving on my thesis committee and helping me produce my thesis into the final product. I would also like to thank Zach Johnston, C.W. Skinner, and Johnathan Bolton for their individual input and excel help. Finally, when finishing in time looked like an impossible feat, Katheryn Skinner kept me positive and supported me until the end.

Table of Contents

List of Tables

List of Illustrations

CHAPTER 1 INTRODUCTION

Manipulating other individuals out of confidential information or property is a common thieving technique. Known generically as a con, social engineering includes the techniques of using social and psychological manipulation to complete the objective of getting others to complete actions. ‘Social engineer’ is the term given to a person who uses techniques of persuasion and manipulation to cause others to divulge private information or complete actions that may not ultimately be in their best interest (Hadnagy, 2011). Social engineering involves extensive knowledge of the human mind and social interactions in regards to how people think, handle, and react to social situations. Successfully manipulating and persuading other humans requires knowing how individuals will interpret and react to certain situations.

Social engineering requires a combination of skills to exploit humans. These exploits can happen in both remote and physical social interactions where attackers approach the target. In order for social engineering to be successful, there needs to be a clear presence of social interaction between the two parties. Remotely, social engineers develop a tangible interaction, which is “giving physical form to digital information” (Hornecker and Buur, 2006, p. 2).

Humans can be seen as the weakest link in information security systems. By influencing humans, social engineers can extract confidential information or further their own interests (Huber, Kowalski, Nohlberg, and Tjoa, 2009). No matter how advanced the technical security system, companies can usually be infiltrated through the human element with little effort. In essence, social engineering is not a problem with technical security but one of human security.

Social engineering is typically seen as the process of extracting confidential information from a target. By using skills of influence, persuasion, and manipulation a person can convince someone to complete actions they want. People can be influenced by a person's body language, voice tone, physical appearance, and how the person words their sentence. In some cases, a victim will never realize social engineers targeted them. Individuals who use social engineering techniques rely on their social interactions with others to achieve access to private information. A typical American citizen interacts with other individuals on a daily basis and also on many different levels. Some of the individual's daily interactions are casual with family or friends, while other interactions are professional. Employing skills and knowledge of body language, dialect, tone, and facial expressions is vital for successful social engineering.

CHAPTER 2 LITERATURE REVIEW

2.1 Historical Context

In 1894, the term 'social engineering' was coined by Dutch industrialist Jacob C. van Marken. The term was used to describe how industrial plants needed to manage their employees, as well as the machines they worked (Conheady, 2014). Although 'social engineering' is a newer term, the act of persuading and tricking individuals into doing actions that may not be in their best interest has been around as long as societal interactions have existed. Perhaps the first well-known documentation of a social engineering act is the 'Trojan Horse' from Homer's *Iliad* in 1240 BC (Thompson, 2013). The very penning of this famous epic only showed that humans of that time were very much aware of conducting frauds on their fellow human beings. With the evolution of technology, humans began to perfect how they targeted individuals to maximize success of manipulation. Today, social engineers have adapted their skills to involve technology, which is quickly growing in its functions and use. As a modern salute to the *Iliad*, certain computer system infections were given the name ‘trojan virus’ (Deshpande, n.d.). These computer viruses are embedded secretly to an attractive link or download a victim opens. The virus hides in the victim’s computer and will even send information back to attackers without the victim’s knowledge. The virus can also remotely install malicious programs onto the infected computer.

In previous decades, famous tricksters such as Charles Ponzi and Frank Abagnale were noted for their use of social engineering skills in physical social engineering frauds. Ponzie and Abagnale realized the potential personal benefits of manipulating other individuals simply by how they presented themselves. In the 1920s, Charles Ponzi used his manipulation techniques to trick investors. Similar investment schemes today are often referred to as a ‘ponzi schemes’. The scheme involves fraudulently investing money and paying previous investors not from profit but from new capital from new investors. Around the 1960s, Abagnale was known for his ability to manipulate and deceive individuals. Through his young adult years, he held over a dozen different identities (Conheady, 2014). Today, social engineers are adopting new methods to target and manipulate victims through technological means. The attacks can never be completely eradicated but actions can be taken to predict and prevent malicious social engineering.

2.2 Remote Social Engineering Attacks

Social engineering in past decades was mainly an issue arising from face-to-face interactions of attackers and selected targets. With the exponential rise of technology and the

Internet, social engineers are expanding their craft and employing their skills of influencing to manipulate people out of confidential/privileged information via technology. Approximately 85% of American adults use the Internet on a regular basis (Perrin and Duggan, 2015). Modern social engineers are using that knowledge to target victims remotely. Remote social engineering attacks are done through technological outlets such as email, computer, mail, and phone. Due to the impersonal and quick delivery, remote attacks are preferred over physical to reach multiple targets at once. With the growth in technological communication connecting more people within seconds, malicious social engineers are turning their talents to target victims via cyber means. Remote attacks have similar aspects to physical such as the engineer will need to conduct background research on the target, develop a pretext, and then use the persuasion skills to influence the target to take desired actions. A big difference between the two attack methods is that remote attacks do not require approaching the target in person, which helps them if they lack personal communication skills (Conheady, 2014). Another difference is that remote attacks allow social engineers to complete their attacks on more flexible time lines and to reach multiple potential victims at once, such as with phishing emails.

Phishing emails are estimated to comprise up to 90 percent of the 300 billion emails sent each day (Hadnagy and Fincher, 2015). A phishing "is the practice of sending e-mails that appear to be from reputable sources with the goal of influencing or gaining personal information" (Hadnagy, 2015). Across the Internet, phishing can take many forms such as email/spam, instant messaging, malware phishing, and phone phishing (Phishing Techniques, n.d.). Phishing is essentially the act of tricking computer users to disclose any personal information such as credit card numbers, user account details of password and user names, and other important credentials. Advanced phishing emails can be crafted to appeal to the basic human emotions of greed, fear, and respect for authority, compassion, curiosity, and desire to connect. Phishing emails are mainly sent generically to a large number of randomly selected recipients. However, other phishing emails can be sent more selectively known as spear phishing. Spear phishing emails are those that are sent to a specific target. If a social engineer knows a specific person they want to extract information from, then they will use a spear phish which can be tailored to the targets interests. Though not exactly a common branch of phishing emails, scambaiting occurs when the target starts to scam the original scammer in return by pretending to fall for the email or offering one of their own. (Linniger, 2005). Many phishing emails require the victim to follow a hyperlink. These hyperlinks appear as a convenience to recipients. Hyperlinks however are one of the easiest ways that social engineers can trick recipients of phishing emails. Other phishing emails contain attachments that, when opened, will download malicious software to the user's system (Dmello, Mhatre, Lopes, & Pen, 2013). In 2013, of the 183 billion emails sent daily, it was estimated that approximately 6 billion contained malicious attachments (Tran, 2013).

The European Network and Information Security Agency developed a list of four key components to help individuals avoid falling for a remote social engineering attack (Conheady, 2014). The first component recipients should verify is the legitimacy of the request in the email. The request and instructions in an email should be questioned as to whether it is legitimate and a typical request the users would receive. If the recipient is even slightly questioning legitimacy, they should contact the company directly or inform their IT department. The second component is the information that the email is asking be provided. The value and significance of the information requested should be questioned as to how it could potentially be used if given. Many phishing emails ask for credit card information or passwords, which should be only be given when the user is sure of the sender's legitimacy. The third component is the source of the request and whether it is a genuine source. If the recipient is unsure of the sender, they should investigate via the Internet to see if other users have reported the same phish. If the source cannot be verified as a genuine source, the recipient should report the email and not follow along with the email's request. The fourth component is the timing of the email. Many phishing emails are sent at interesting times, such as early in the morning or during certain months. Phishing and emails that are sent to users from foreign countries might be sent at unusual hours because of time zone differences (Conheady, 2014). If email account users question each email they receive with these four component, they can drastically reduce their chances of falling for a phishing email.

2.3 Techniques of Persuasion and Influence

A social engineering attack is comprised of distinct components that are important in determining success: research, pretext, and rapport. Conducting extensive research into the target is an essential step in forming a believable pretext and a strong rapport. A lack of research can make attackers more hesitant in answering questions or appear uncomfortable to the target, which can raise concerns in the target (Hogan and Speakman, 2006). When entering into the first phases of completing a social engineering mission, attackers will perform research on targets to learn valuable information through means of Internet searches or physical surveillance. In regards to attacking a single individual, detailed notes should be recorded on aspects of the target such as personal interest, employer, life history records, social media/internet presence, and daily routine. When the target is larger, such as a corporation, research should be conducted on the physical aspects of the building, employees' personal information, facility companies employed, clients, past history, and known company procedures (Conheady, 2014). By having a basic background on the target, conversations can flow easily and social engineers can be prepared for any surprise questions that could arise challenging the legitimacy of the attacker's request. This research will allow the attacker to formulate the best plan on how to approach the target with the greatest chance of not being discovered as a phish. When selecting a target, social engineers first need to evaluate what their end goal is. If they are simply trying to attack as many people as possible, the targets will be chosen rather randomly as they obtain the contact information such as email addresses. However, if the target is more precise, such as a large technology corporation, social engineers will study the employee base to select targets. Depending on the manipulation used, social engineers will select targets that appear to be most susceptible and unaware of the dangers of potential breeches in security (Hadnagy, 2011).

Formulating a convincing pretext is the next important component of the social engineering attack and should be handled delicately. In order to make the conversation more authentic, social engineers will develop a role, much like being an actor, who has a better chance of convincing the target. The process is called developing a pretext, a term "defined as the act of creating an invested scenario to persuade a targeted victim to release information or perform some action" (Hadnagy, 2011, p.78). A successful pretext cannot be completed without the research on the target. Social engineers rely heavily on pretexting skills in order to successfully complete their social engineering mission by understanding how to best connect with the target. To help the victim become more likely to believe social engineers, a created pretext needs to feel realistic to the target. Saying one wrong thing or behaving in an off-putting manner can cause the target to shut down and essentially end the chance to successfully get information.

A pretext needs to be simply designed so that it can be easily remembered. Creating a complex story with many different entities involved only opens opportunity for mistakes and loss of rapport. Instead, social engineers will create a pretext that will be simple for them to remember and recall. A pretext is basically acting a different role and the more different the role is, the more difficult it will be to present it convincingly (Ivaturi and Janczewski, 2013). To make the pretext appear more believable, social engineers will try to include their own interest or natural talents. If the pretext contains too many aspects out of their comfort zone, they will likely not be able to project it accurately. In order to execute a successful pretext scenario, social engineers also need to keep personal emotions from interfering. The scenario created with the pretext should be treated as its own entity. If personal emotions show through, they are likely to be conflicting with the pretext. No matter attackers' personal moral values toward a certain topic, they need to remember to act like the pretext would act. If attackers allow their own values to interfere, they could show conflicting personas to the victim. Researching the target's dialect and accent can help social engineers be accepted by the target. Dialects and accents are unique to different regions of the world. If a foreign dialect is detected, it can cause a person to form instant impressions about the attacker, which can be negative. Anyone with a different accent could cause the victim to become untrustworthy or think they are of lower intellect (Hadnagy, 2011).

The underlining key to a successful social engineering attack, particularly in physical attacks, is the strength of the rapport with the victim. Rapport is built when there is “connection with someone and putting him or her at ease” (Hadnagy, 2011, p.170). This bond promotes trust and the desire to be as helpful as possible. It also ensures a strong connection between the two parties that allows smooth understanding and communication. Forming a rapport with the target should be one of the first goals the attack completes. A good rapport creates an atmosphere of trust with the target. The perceived trust causes them to be more likely to comply with the attacker’s request. Rapport is a very fragile aspect of the attack that can be broken at the slightest suspicion of uncertainty. Once rapport is questioned, social engineers’ chances of success manipulating that specific target significantly decreases.

Principles of sociology and psychology are very relevant to any person wanting to conduct a social engineering fraud. Rapport building is a tricky technique that takes practice and knowledge of how the human mind operates. By using social communication skills of persuasion, manipulation and influence, social engineers will attempt to build rapport with anyone they encounter. Attackers know that if they gain a connection with a victim, particularly one that causes the victim to feel attaching emotions to them, then the success of completing the attack is raised. However, if social engineers cause any sort of uncertainty in the target, it can break the rapport (Hadnagy, 2011). Once any sign of uncertainty is formed in the target, they are essentially useless to social engineers. If social engineers discover a break in rapport, they will typically exit and search for another target.

To address how social engineers create and use rapport to their advantage, it is important to break down the components and analyzing them. In one of his books, FBI veteran Robin Dreeke sums up the top ten techniques to rapport building. The ten most important techniques to remember when building rapport are artificial time constraints, ego suspension, managing expectations, slower rate of speech, sympathy/assistance, accommodating nonverbal, when and why questions, asking how, reciprocal altruism, and quid pro quo (Dreeke and Robin, 2011). These techniques are easily applied to physical social engineering attacks but can also be applied to remote attacks. With a solid understanding of each of these elements, a stable and trusting rapport can be easily built.

Artificial time constraints are put in place by social engineers because the rapport needs to be completed within a certain time period. Creating a rapport with the target should be done as quickly as possible. Building rapport quickly eliminates the changes of the victim becoming questioning or hesitant of the situation. An artificial time constraint should be used as a guideline and should be handled with care to not be forced. If social engineers sense that rapport building is taking too long, they should take care not to appear anxious or rushed (Dreeke and Robin, 2011). Within the creation of both physical and remote social engineering attacks, time spent building a rapport should be done as swiftly as possible. With every passing minute, the victim might be exposed to security education or information that might cause them to break contact.

A good rapport is built around the victim's feelings and is centered on the target at that moment in time (Conheady 2014). Social engineers need to employ ego suspension to keep personal feelings under control. Ego suspension is when social engineers ignore their own ego in attempts to elevate the victim's ego. By suspending ego, attackers can focus more on inflating the target's ego and increasing a quick rapport building. Practicing ego suspension is closely related to how the engineer makes sure to not let personal emotions interfere (Dreeke and Robin, 2011). If social engineers let their own ego effect personal actions, it could cause the victim to become suspicious.

Social engineers also need to manage their expectations of what they want to receive from the encounter with the target. Keeping expectations in relative relation with the information being sought will help social engineers not push too hard on the target. Keeping the demands simple and direct have the least likelihood of raising concerns in the target. If the target becomes untrusting of the encounter, they might close themselves off from trying to connect with social engineers (Hadnagy, 2011). When targets become suspicious, they will immediately shut down any chance of a successful rapport. A suspicious target is essentially useless and offers no further advance for social engineers (Dreeke and Robin, 2011). Social engineers need to be aware of this scenario and slip away to find another target. Suspicion might grow further in the target and cause them to alert others. Burning of a target can happen if not handled correctly so social engineers need to have backup plans in place.

Social engineers need to take care to control their speech when talking with the target in person or on the phone. Talking at controlled rates will make the engineer appear to be comfortable in the situation. People in general tend to speak quicker when they are nervous or trying to hurry something along. When targets hear a person talking rapidly and nervously, they are more likely to become suspicious of what the person is asking them to do. Having controlled communication will help ensure the target feels comfortable in the conversation (Hadnagy, 2011). Creating an illusion of sympathy during the communication with target is another important rapport building technique to master. An easy way to manipulate the target is to create sympathy and appear as if assistance is needed. Most individuals in society will choose to want to help someone who is sad. Playing off the target's moral humanity to want to help or assist people in need or in pain is one of the easiest tricks in a social engineer's book (Dreeke and Robin, 2011). Being stricken with sympathy for someone can cause targets to abandon any reason or logic and cause them to want to help in any way possible. All social engineers need to do is create an illusion that they need help in order to draw targets into doing what they want.

Maintaining nonverbal body actions is very important in any physical social engineering attack. Social engineers need to be practiced in reading emotional cues from body language to be able to steer the conversation in the desired direction. Equally as important, social engineers need to be aware of personal body language and the possibilities that come with employing it to manipulate feelings in the target. A big percentage of face-to-face communication comes from a person's body language. When talking with others, people subconsciously project emotions or feelings through the way they move their body (Hadnagy, 2014). The way the torso faces, the way the hands and feet are placed, and micro expressions that cross the face all provide hints at what the person is truly feeling.

Social engineers need to be well prepared to answer any questions that targets might ask during the interactions. When interacting with another person, being able to ask and also answer the ‘how, when, and why’ is important. Knowing how to ask questions to get the desired answers is the attacker’s key (Dreeke and Robin, 2011). If targets ask a question that social engineers cannot answer, the target could easily shut down any request made. If targets question for any reason, attackers need to be able to think on their feet to respond without raising any suspicion. Also, social engineers need to know what questions they can ask targets in order to get the needed information without overstepping boundaries.

Remote social engineering’s occurrences have become more common as the world becomes increasingly connected in the cyber realm. Today, individuals can shop, talk with others, and pay bills all without leaving their homes thanks to the Internet. An individual can use a password account to handle their financial matters such as paying taxes and handling credit cards. The connectivity that the Internet has brought humanity is revolutionary. However, with all the enjoyment and convenience the Internet brings people, it also brings the bad people. Internet crimes are occurring more often now that so many people use the Internet. Social engineers can turn their manipulation skills to the Internet through frauds where they get individuals to unknowingly hand over confidential information. To steal a person’s identity or money, a social engineer just needs to learn the account’s password (Kirda and Kruegal, 2006). Other confidential information that can be targeted is intellectual property created by academia. Intellectual property is research findings and results discovered by researchers that can be stolen for use by others illegally. Academic institutions across the United States are subject to being targeted for the intellectual property they create.

In face-to-face interactions, social engineers have to be more concerned with how their appearance and presentation is used to convince victims during interactions. However, in the cyber realm social engineers simply have to be able to write convincingly to gain a sense of trust and manipulate individuals. When the recipient feels as if they trust the sender, they will feel compelled to do what the sender is requesting. Pretending to be someone else is considerably easier when social engineers are not facing the victim. It can be argued that social engineering is actually easier in the cyber world (Hadnagy, 2015). Luckily, protecting against social engineering in the cyber realm, such as phishing emails, is easy with simple education on how to recognize it. In order to provide educational tips that can protect academia from the social engineering attacks of phishing emails, patterns of how academic institutions are attacked should be studied and identified.

Finally, social engineers need to focus on reciprocal altruism and quid pro quo to ensure a quick and strong rapport with the target. Reciprocal altruism is a technique when social engineers rely on the concept that when they give targets something, targets will then feel obligated to reciprocate and give something in return (Rusch, 1999). A basic common aspect of human nature is that humans will feel more inclined to help those who have helped or given them something first. This phenomenon is commonly known in the Sociological field as the Social Exchange Theory.

2.4 Social Exchange Theory

Social Exchange Theory (SET) is a paradigm of tenets and assumptions that offer explanations to how and why individuals act the way they do in face-to-face interactions with one another. In other words, when individuals in society interact, if the first individual gives/offers something, the recipient will feel compelled to reciprocate. SET broadly describes "actions that are contingent on rewarding reactions of others" (Blau, 1964). The theory explains the way that individuals exchange and negotiate between each other during their social interactions. Social Exchange Theory as a paradigm was created in the late 1950s to 1960s but the theoretical ideas and roots first began as early as the 1920s (Emerson, 1976).

Roots for SET appeared heavily in the field of Psychology through the work of B.F. Skinner who was a psychologist and a behaviorist in the 1900s. Skinner focused on the actions of individuals in order to discover why humans behave the way they do. His work showed that actions were dependent on any consequences that were experienced in previous actions. Through Skinner's process of operant conditioning, he discovered that having rewards and punishments as consequences for the first action would affect future actions. By using laboratory animals in his experiments, Skinner observed animals that once an animal received a certain rewards, it would try to recreate its actions again in hopes of receiving the reward again (Skinner, 1965). The ideas of expecting a certain reward for actions that Skinner discovered have a link to the tenets of SET that Homans applied and expanded on. If an individual received a positive reward after completing an action, then they were more likely to repeat that action in hopes of the same reward.

Social Exchange Theory operates in part under the assumption that individuals experience a type of cost during exchanges in order to reap the future rewards. Sociologist George Homans was a major contributor to the SET paradigm with the concepts of cost/rewards and the summarizations of human behaviors. Homans theorized that people view rewards and costs of their actions in very distinct ways. He believed that if a result of an action brought value than the individual is more likely to do the action again (Homans, 1958). Cost refers to anything that an individual has to give up in order to receive a certain outcome and can result to a decrease in the likelihood of every doing the action again. Homans determined that every social exchange has transference of costs in order to gain rewards. SET assumes that individuals are motivated by self-interest. If an individual is given an opportunity to gain something they want, chances are they will agree to the conditions of the exchange (Homans, 1958). There might be a cost involved at first, but the individual rationalizes the cost if the rewards are even greater, so in the end they appear to end up better off then they started.

Homans also contributed to the SET paradigm with the development of three basic propositions that describe human behaviors. Homans believed that given the circumstance of the situation, the individuals would behave a certain way in response. The first is referred to as the 'Success Proposition', which explains that the more a certain action is rewarded, the more likely the individual is to do that action again (Cook and Rice, 2013). The second is the 'Stimulus Proposition' that states any stimulus from an action that was rewarded will mean similar stimuli in the future will cause similar actions. The third proposition, the 'Deprivation-Satiation Proposition', states that if an individual receives a certain reward multiple times, the reward is then less valuable in the future (Emerson, 1976). These propositions are underlining factors in how individuals will behave when interacting with others. Not only will knowing about the propositions expand an individual's knowledge of human behavior, but they can also learn to use it to their advantage. With an understanding of Homans' propositions, a person can manipulate the exchange process with others to go in their favor, much like social engineers do.

Another important assumption of the Social Exchange Theory is that humans are rational beings and will want to participle in an exchange and reciprocate if they are being offered something of value. One of the main contributors, sociologist Peter Blau, believed that social interactions are created by a reciprocal exchange of tangible and intangible things. In SET, there is a basic exchange rule that many social exchanges are regulated by a norm of reciprocity. Reciprocity is a repayment or exchange of an object/favor in return for an object/favor. A reciprocal exchange is understood as one that does not include explicit bargaining; rather one party's actions are contingent on the other's behavior (Cropanzano and Mitchell, 2007, p. 876). If the first party offers some help to the second, then the likelihood of the second person offering help in the future is greater since they will want to repay the metaphorical debt. Social exchange is a central piece of the social lives of individuals and is an underlying component of relations between groups (Cook, 2003). The reciprocity may appear to be given at a small cost to the individual but the reward/favor's benefits they are receiving will convince them the exchange is appropriate.

Whether social engineers are aware of it or not, they rely heavily on the Social Exchange Theory's principles to be successful in attacking people. The core tenet of social engineering is that manipulating people is the key to obtaining confidential information. With a solid understanding of the principles and assumptions involved in the Social Exchange Theory, social engineers can cater their actions and language to easily create a need of reciprocation/response in their victims. Social engineers can use a cost-benefit mentality when they are interacting with others by trusting that if they give something to an individual that that person now is indebted to return the favor. If social engineers create a scenario where they either offer some sort of assistance or create the illusion that they did, then targets will feel gratitude and offer reciprocation. Social engineers may word their request in such a way that recipients may have a cost of having to provide something, such as information, but will hope to get a reward in return that will make it worth it. Social engineers can also approach with a proposition that if targets do something for them then social engineers will reciprocate with an even bigger favor (Emerson, 1976). Humans are more likely to accept a situation or task when they stand to gain something in return from the transaction. Social engineers have to make sure the social exchange is clearly in the email's contents so the recipient will be likely to respond.

CHAPTER 3 METHODS

3.1 Research Strategy

Social engineering is a method of targeting victims that is easy to administer with little cost and consequences to attackers. The growth of technology and the Internet has made remote social engineering techniques even easier to administer. Phishing and spam emails are the most common form of remote social engineering attacks. Whether social engineers are aware, they are using the principles of Social Exchange Theory when sending phishing email attacks. In order to increase the likelihood of having phishing emails seen, the emails are written to make the user feel indebted to the sender in some way. This feeling of debt can be gratitude from a special offer or simply feeling like the sender has already done something for them so they should return the favor. A recipient will feel the need to open an email that is asking for assistance and in return rewarding the recipient for their trouble. Or, the recipient might feel indebted to the sender if the sender seems to have helped them in any way, such as appearing to have protected their accounts or taking the time to inform them of possible problems. This perceived help might cause the recipient to want to reciprocate back and comply with what is being asked. Big organizations like academic institutions are seeing thousands of phishing attacks sent to users daily. By employing the underlying principles of Social Exchange Theory, social engineers are convincing individuals to provide confidential information.

The present study analyzes a phishing email data sample from a participating academic institution. The amounts of previous research studies surrounding this specific topic are few and rare. Focusing on academic phishing emails is rather exploratory and exact results are hard to predict. However, the results aim to identify patterns of university recipients being targeted, common subject matter and emotions used to target individuals, and written structure typical of phishing emails. The results will provide universities and others with knowledge and insight on the craftsmanship behind phishing emails that will aid them in creating security policies to counteract remote social engineering attacks.

The study seeks to answer the three following research questions:

1: Which days and months are academia email account users most likely to receive phishing emails?

2: Are account users in certain job positions more likely to received phishing emails then others?

3: Do phishing emails have a general pattern that can be studied for future education awareness?

3.2 Data

The goal of this study is to understand what makes phishing emails appealing to recipients and what types of university employees are most likely to be targeted. The data in this study is a sample of actual phishing email attacks on academia from a participating southern university's information technology department. The collection period covers a span of seven months, from June 2015 to December 2015. The majority of phishing/spam emails sent to the university email users are stopped by the IT's spam filters. However, a few phishing emails each day get past filters and are sent to users. Recipients who recognize the phishing attempt report them and the emails get in the system to an IT employee. The IT employee then forwards all phishing emails from recipients to an email inbox for study collection.

The big data program Splunk then searches for data in each of the emails. Splunk is a big data program that makes the collection, storage, retrieval, and analysis of large amounts of data easier to manage. Splunk specializes in the recording all activities and behaviors of machine data that includes the users, servers, and network's actions. The machine data Splunk handles also "includes configurations, data from APIs, message queues, change events, the output of diagnostic commands…and sensor data from industrial systems" (Splunk, n.d.). The university's email software sends tracking logs to the Splunk servers for all emails that pass through employed individual's accounts.

The title of each email is placed into a search window on Splunk to find the relevant recipient and time data. Splunk then retrieves all the relevant data for that email that from the system. Relevant data of interest is the start time the email is first sent to a university recipient, the stop time when the IT department blocks the email from the system, and data on every recipient who received the email. Splunk did not collect university student emails unless that student also has a university job. Student email account data is kept on separate technology servers. The only student recipients that this study collects are in the university system as an employee. Splunk could not provide information on which if any recipients opens the phishing emails they receive.

Since the study focuses on the phishing emails and not the human recipients themselves, the data is stripped of any identifying information (name, university ID, and email address). During the seven-month collection period, a total of 240 phishing emails were successfully searched in Splunk and the relevant recipient data are collected. For each phishing email in the sample, the raw data is extracted from Splunk and placed into Excel tables for organization. After the phishing data are organized in Excel, the data is then moved to SPSS for statistical testing.

Through the use of the statistical program SPSS, the data is run through different testing methods for quantitative results. Descriptive statistics are used to help show the distribution of the different variables. Then, bivariate correlations are used to determine any significant relationships among variables. A Value of 1 for a Pearson Correlation test shows a strong correlation. If the Pearson Correlation value is close to 0 it indicates a weak relationship. Finally, multinomial logistic regression tests were used to compare the each of the dependent variables with the three independent variables of recipient information. A multinomial logistic regression test classifies a regression that is used to predict possible relationship outcomes with a dependent variable that is categorically classified with independent variables (Statistics Help for Students, 2008). For each mulinomal logistic regression used in the study, the largest category of the dependent variable was chosen as the reference group.

The author analyzed content patterns and themes in each email to distinguish any common patterns of structure. In order to do so, a screenshot of each phishing email was collected and placed into a corresponding word document. Every time a noticeable feature was distinguished, it is recorded and the final results are totaled. In order to keep recipients anonymous, no names or university identification of recipients are collected in the screenshots. The presentation and grammatical composition of each phishing email is analyzed to determine factors that are common among phishing emails. The heading and body of the phishing email in particular is studied to determine any similar characteristics between emails. The phishing emails' subject is analyzed to distinguish any use of persuasion techniques and methods of appeal to the victim. Common indictors are grammar and spelling errors, non-descriptive content information, and broad greetings.

3.3 Variables

The study defines two dependent variables of common subject and emotion categories. The study organizes the emails into common related subject categories and common emotions the emails are trying to provoke. The first dependent variable, “Subject Emails’” classifies each of the 240 emails as one of five types of subject categories. Grouping the 240 emails into five subject categories reduces the data to be more manageable for testing. The five subject groups are chosen because the phishing email samples content tends to focus around the five areas of “Financial”, “Helpdesk IT”, “Login Password”, “Miscellaneous”, and “Webserver Mail”. “Financial” and covers any emails that had a subject line referring to financial matters such as bank accounts. The second group “Helpdesk IT” is the classification for phishing emails that have a subject/title referring to IT (information technology) topics or helpdesk-type request. “Login Passwords” is the classification for any of the emails that have a subject line referring to password reset or account login troubles. “Webserver Mail” includes any emails that deal with email related subject matter, such as a webserver error. “Miscellaneous” covers any subject that does not fit within the other four categories.

The second dependent variable classifies the 240 emails into emotions. The emotions chosen are ones that the title aims to invoke from the recipient. The emotion that is first triggered by reading the subject line can greatly impact whether the individual will open the email or delete it. The 240 emails were grouped in the four emotion categories of “Anxious”, “Confused”, “Curious”, and “Greed”. “Anxious” is assigned to any email that will make the receiver anxious and nervous upon reading the subject line in their inbox. “Confused” is any email with a subject line that causes the receiver to feel confused about why they would be receiving the email. “Curious” are the emails that make the receiver want to open it because of intrigue and curiosity about the contents. “Greed” is the classification for any of the 240-phishing emails that try to create a feeling of want and greed from recipients.

For each phishing email, three important time components of “Day Sent”, “Month Sent”, and ‘”Minute Duration” are documented as independent variables. “Day Sent” is the exact day of the week that each email was first sent to a university recipient. The “Day Sent” variable is coded as the seven days of the week (1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday, 6=Saturday, and 7=Sunday). “Month Sent” shows the month that each phishing email was sent to users. The “Month Sent” variable is coded as the calendar months of June through December due to the collection period time frame (6=June, 7=July, 8=August, 9=September, 10=October, 11=November, and 12=December). “Minute Duration” shows the total in minutes that each of the 240-phishing emails was active in the university email system. For each email, the “Minute Duration” starts when the phishing email is first received by a system user’s account. The “Minute Duration” stops being calculated when the IT department blocks the email from being sent to further university account users.

To determine which university recipients might be most at risk to receiving phishing attacks, the study analyzes the independent variables of the role of the recipient, the recipient’s university title, and the recipient’s division. The details of the recipient only provided organization information for how they were employed with the university. In other words, all of the titles and classifications are taken from what the university used to label an employee’s job description. The “Role” variable distinguishes if the recipient is a student, employee, or retired. The “Title” variable is the recipient’s individual job description. The “Division” variable is the university department the individual is employed under.

The 240 emails’ recipient varaibles are coded to reduce the data in an easier format for the statistics test as well as offer more details into who is mostly to be attacked with phishing emails in academia. The “Role’” variable is coded using five categories to group together like values. The five “Role” categories used in labeling the data are “1=Employee”, “2=Affiliate”, “3=Department”, “4=Retired”, and “5=Student”. The variable of recipient’s “Title” was coded as “1=Faculty” or “2=Non-Faculty”. The “Faculty” classification combines like employments of the recipients listed such as a professor, lecture, teacher, etc. All other recipients are given the label of “Non-Faculty”. Finally, the variable of ‘Division’ is coded using five categories of “1=College”, “2=Administration”, “3=Library”, “4=Affiliated”, and “5=Extension”. The five categories for division help break down the different recipient’s job position areas in the university.

CHAPTER 4 FINDINGS

4.1 Content Theme Analysis

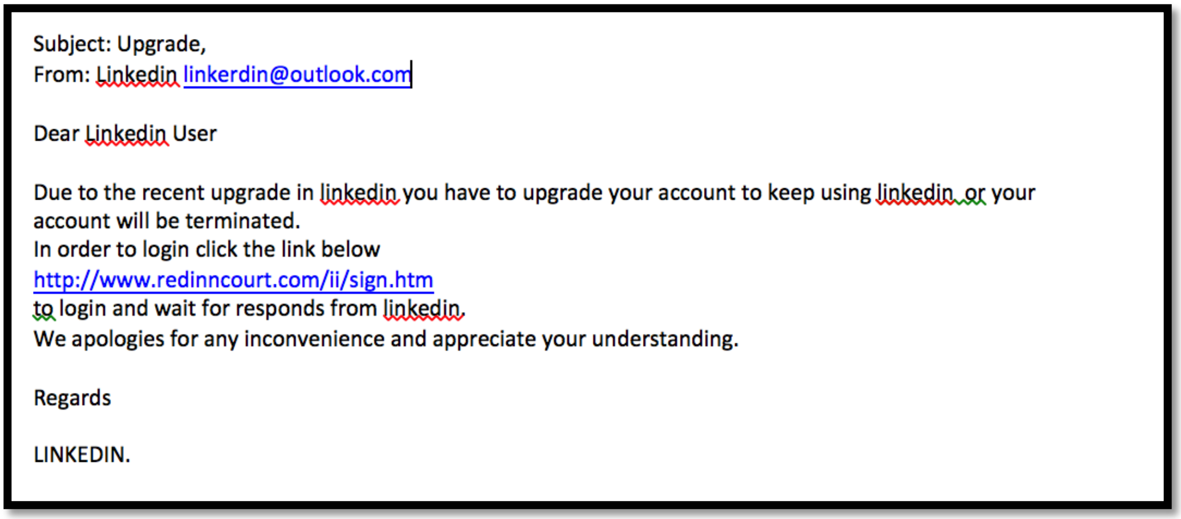
The phishing email’s subject titles and content were individually analyzed for content themes. For this portion of the study, each of the 240-phishing emails was analyzed individually to access any features that might indicate a phishing attempt. Many important observations from content theme analysis should be noted. The three main observation areas were presentation and body composition. Awareness of content analysis of phishing emails can help increase recipient's chances of spotting fraudulent emails.

From the data sample, phishing emails tended to have some reoccurring themes in terms of how they were constructed. When looking at an email, the recipient needs to be aware of the presentation components of the email, which can over apparent indicators of the email being fraudulent. After reading through the 240-phishing emails, compositions components that were recorded as significant were impersonal greetings/addressing and use of extra features of hyperlinks, logos, and attachments.

The first notable presentation component was impersonal greetings and addressing/tone of the emails. Once opened, an email has to keep the target's attention by continuing to sound convincing and legitimate. However, the recipient should make note of how the opening and closing of the email are written. 111 of the 240-phishing emails contained an impersonal greeting. These 111 greetings addressed the recipient broadly such as 'email user' or 'account user'. If the email kept a generic greeting and does not address the recipient directly, it could be a phishing email sent to many recipients at once. Throughout the body of an email, the recipients should be alert for how they are being addressed. If an email used words that are very broadly referring to them as 'customer' or 'friend' repeatedly, the recipient should proceed to continue critically. The closing statement as well will tend to be more generic when it is part of a phishing campaign. This way, social engineers are able to send the same email to a mass number of targets without having to personalize them each time.

The second presentation component was the use of extras features to the email. Hyperlinks were the greatest method used by social engineers to try to enter systems. If clicked, a hyperlink could either instantly download malicious content to a user's computer or can lead the user to a fraudulent site to enter PPI that can be collected by social engineers. Hyperlinks appeared in 216 of the 240 emails. As seen in Illustration 1 below, one of the emails from the sample that displayed an impersonal greeting as well as a suspicious hyperlink. As seen in Illustration 1, the sender address is spelled "linkerdin", which added an extra letter to the LinkedIn name. Although the added letter was a slight detail, recipients should make sure they spend the extra seconds to examine it before they clicks on any of the link. Hyperlinks can be so damaging to the security of an academic system but at the same time can be so easily protected against. Account users on the university system should be informed of how prevalent malicious hyperlinks are and how damaging simply clicking one can be. If the user thinks a hyperlink might take them to a site they know, they can always use the 'hover' method before clicking. By hovering their mouse curser over the link, typically the computer will show a drop-box of the website they will go to. This can show users if the website is one they recognize or not.

Illustration 1: Suspicious Hyperlink and Impersonal Greeting Email



Subject: Upgrade,
From: Linkedin linkerdin@outlook.com

Dear Linkedin User

Due to the recent upgrade in linkedin you have to upgrade your account to keep using linkedin. or your account will be terminated.
In order to login click the link below
http://www.redinncourt.com/ii/sign.htm
to login and wait for responds from linkedin.
We apologies for any inconvenience and appreciate your understanding.

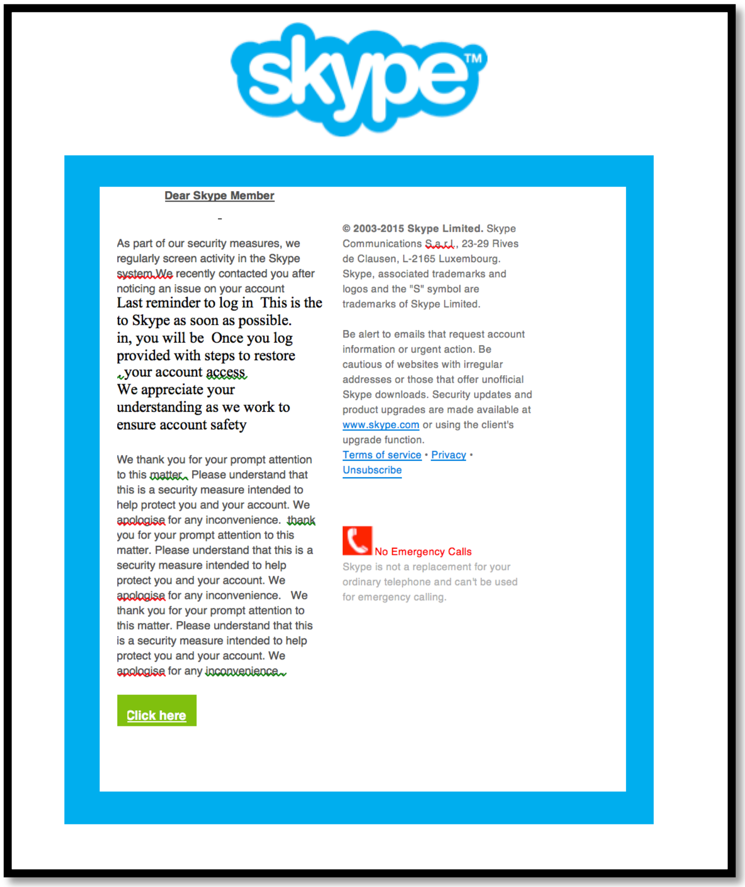
Regards

LINKEDIN.

Logos were found in a number of the phishing emails. 17 of the 240-phishing emails used a company logo to convince the recipient it could be a real email. Many large corporations have their own logo that they prevalently display on any emails that are sent to their clients. Social engineers use logos in order to make their fake emails appear legitimate. By using a company logo that a recipient will recognize, there is an implied level of trust in the recipient that the company is one they know. This level of trust the recipient has for the real company will lead them to be more likely to click on any links in order to do their part in helping the company with the request, which displays underlying themes of Social Exchange Theory. Social engineers can easily copy and paste the exact image a company uses so the recipients cannot tell if it is legitimate. Even if an email has an official logo, a recipient still needs to critically examine the rest of the email for any obvious indicators of phishing email. Illustration 2 from the sample was an example from the sample of using a logo to convince recipients of legitimacy. The email even displayed a border and format to match the logo that can help convince uses it is really from the

Skype corporation. Attachments are also used in phishing email attacks but not as often as hyperlinks and logos. In the sample only 3 emails had attachments but should still be mentioned. Just like hyperlinks, attachments can deliver viruses and malware to a user's computer if opened. The attachment can be anything from a PDF to a picture. Recipients should never open an attachment emailed to them unless they are expecting an attachment from a known sender.

Illustration 2: Use of Logo in Phishing Email Example



Dear Skype Member

-

As part of our security measures, we regularly screen activity in the Skype system.We recently contacted you after noticing an issue on your account

Last reminder to log in This is the
to Skype as soon as possible.
in, you will be Once you log
provided with steps to restore
.your account access.
We appreciate your
understanding as we work to
ensure account safety

We thank you for your prompt attention to this matter. Please understand that this is a security measure intended to help protect you and your account. We apologise for any inconvenience. thank you for your prompt attention to this matter. Please understand that this is a security measure intended to help protect you and your account. We apologise for any inconvenience.   We thank you for your prompt attention to this matter. Please understand that this is a security measure intended to help protect you and your account. We apologise for any inconvenience.

Click here

© 2003-2015 Skype Limited. Skype Communications S.a.r.l., 23-29 Rives de Clausen, L-2165 Luxembourg. Skype, associated trademarks and logos and the "S" symbol are trademarks of Skype Limited.

Be alert to emails that request account information or urgent action. Be cautious of websites with irregular addresses or those that offer unofficial Skype downloads. Security updates and product upgrades are made available at www.skype.com or using the client's upgrade function.

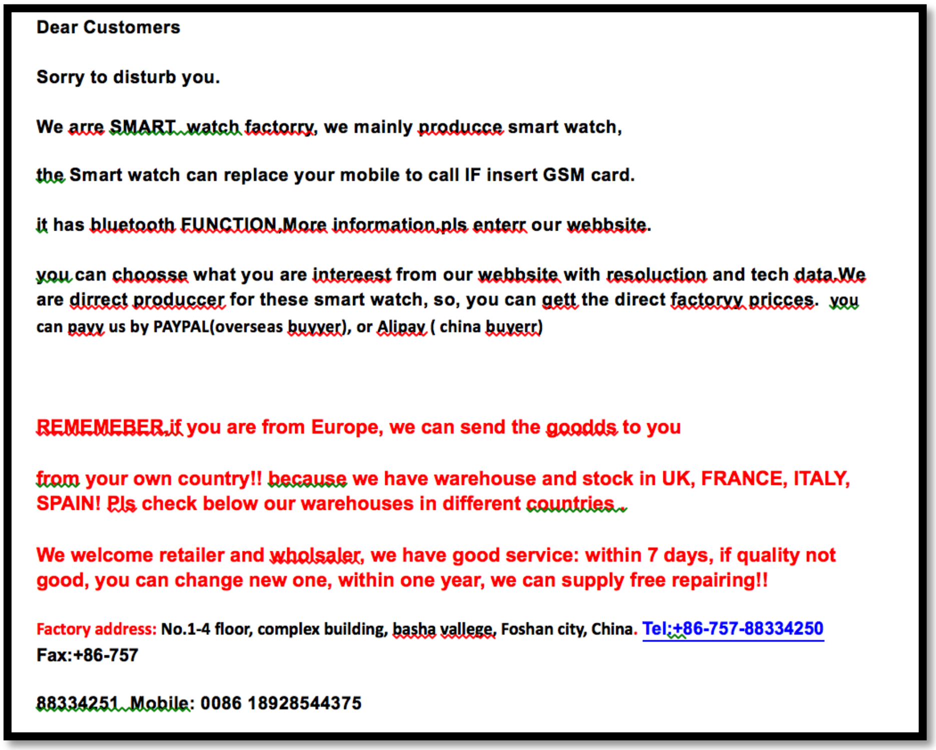
Terms of service • Privacy • Unsubscribe

☎ No Emergency Calls

Skype is not a replacement for your ordinary telephone and can't be used for emergency calling.

The composition factors of each phishing email's body were analyzed for sentence structure and grammar. The easiest structure indicators of a phishing email for recipients to notice are spelling errors. 28 of the collect phishing emails contained spelling errors. Illustration 3 shows grammar issues as well as numerous spelling mistakes found in one of the emails studied. Most of the 28-phishing emails that had only minor spelling mistakes, but the one seen in Illustration 3 was littered with errors. The sentence structure as well in the example was unnatural. 172 of the emails displayed sentences that were structured in a rather unnatural and illogical ways.

Illustration 3: Structure Errors of Grammar and Spelling Email



Dear Customers

Sorry to disturb you.

We arre SMART watch factorry, we mainly producce smart watch,

the Smart watch can replace your mobile to call IF insert GSM card.

it has bluetooth FUNCTION.More information.pls enterr our webbsite.

you can choosse what you are intereest from our webbsite with resolution and tech data.We are dirrect produccer for these smart watch, so, you can gett the direct factoryy pricces. you can payy us by PAYPAL(overseas buyyer), or Alipay (china buyerr)

REMEMEBER.if you are from Europe, we can send the gooddss to you

from your own country!! because we have warehouse and stock in UK, FRANCE, ITALY, SPAIN! Pls check below our warehouses in different countries .

We welcome retailer and wholsaler, we have good service: within 7 days, if quality not good, you can change new one, within one year, we can supply free repairing!!

Factory address: No.1-4 floor, complex building, basha vallege, Foshan city, China. Tel:+86-757-88334250
Fax:+86-757

88334251 Mobile: 0086 18928544375

4.2 Quantitative Analysis

To see the scope of the distribution of the different variables, descriptive statistics were used. The day that the phishing email is an important variable to study because it might potentially show when the university system was attacked most by phishing emails. The frequency distribution of the "Day Sent" that each email first entered into the university email system in seen in Table 1. During the 7-month collection period, 62 phishing emails were sent on Monday, which was the highest of all of the days of the week. The beginning of typical business workweek starts on Monday, which could explain why so many of the emails were sent on Monday. Wednesday and Friday also showed a high number of incidences across the collection period. Being that Wednesday was the middle of the workweek and Friday was the end shows a pattern of when social engineers choose to send phishing emails. Thursday had 22 incidences while Tuesday had 18. Sunday and Saturday had the lowest number emails sent to users. While both of these were days of business operation, they each had very low incidences compared to Monday, Wednesday, and Friday. Of the 240 emails only 22 were sent on Saturday and Sunday, which were days that the university does not hold business hours. Account users will spend the most time in their email accounts when they are working. Social engineers want to send their emails at the times they are most likely to be seen, typically on weekdays.

Table 1: Day Sent

Day	Frequency	Percent
Monday	62	25.8
Tuesday	18	7.5
Wednesday	61	25.4
Thursday	22	9.2
Friday	56	23.3
Saturday	10	4.2
Sunday	11	4.6
Total	240	100.0

Similar to how the day sent was important, the month that the email was sent is equally as important. The month phishing emails were sent can help IT prepare security policies and know when to be alert for the most attacks. The collection period started before the semester started and ended a few weeks after the conclusion of the semester. Seen in Table 2, phishing emails were evenly distributed during the months of July through October. This even and steady flow of phishing emails could be from social engineers know that users will most likely access their university email accounts during the major part of the Fall academic semester. October, however, was when 48 of the 240 emails were sent to users in the university system. The months of June and December had the lowest incidences of all seven of the months. June is a month when the university was open and holding Summer classes, however there were significantly less users on the email servers during the Summer. The phishing email flow into the university started to lower in November and becomes very low in December. The academic semester ends around the first week of December and then employees were given holidays a good portion of the month. When the holiday times of November and December come around, social engineers might send fewer emails because they know that email users tend to check their work email less.

Table 2: Month Sent

Month	Frequency	Percent
June	17	7.1
July	45	18.8
August	43	17.9
September	46	19.2
October	48	20.0
November	31	12.9
December	10	4.2
Total	240	100.0

The third time variable extracted from the data sample, “Minute Duration”, shows the total duration in minutes that each phishing email was active in the university servers. The exact minute duration the phishing emails were in the system, from the first moment they were sent to recipients to the minute the system blocked it, offers important security information. Table 3 shows the total minutes that each of the dependent variable “Subject Theme” five categories emails was in the system. In total, all of the phishing emails with subjects related to “Webserver” stayed in the system 866,433 minutes. “Webserver” emails were also sent to university users 83 times during the collection period, as seen in Table 6, which can equate to why it has the longest total amount of minutes in the system. The “Financial” phishing emails stayed in the system the least amount of minutes with only 98,828 total minutes. Although there appears to be a drastic difference between “Webserver” emails and “Financial” it was important to remember that there were only nine “Financial” phishing emails collected while there were 83 “Webserver”. Proportionally to the amount of cases for each of the five subjects, the “Minute Duration” was rather constant for all of the phishing. Table 3 can be used to learn which types of phishing email take the university system longer to catch.

Table 3: Subject Email Minute Duration

Subject	Frequency
Financial	98,828
Helpdesk IT	592,658
Login Password	515,406
Miscellaneous	398,964
Webserver	866,433
Total	2,472,289

The study examined the actual distribution of the dependent variables. First, tests were run to determine if any of the 5 subject categories that the phishing emails were grouped into were sent to university users more prevalently that other. By using a descriptive statistics, Table 6 shows the total number phishing email occurrences collected for each of the subjects groups. Emails that had message content related to “Webserver” topics showed having 83 incidences recorded during the collection period. This was the highest number of all five of the subject categories meaning that the university recipients are more likely to receive webserver phishing emails than other phishing email subjects. The subject category that had the least number of phishing emails received during the seven-month period was the “Financial” related emails with 9 emails over all. The categories of “Helpdesk IT” and “Login Password” were shown to have high number of occurrences. Table 6 provides a summary of the types of topics that phishing emails focus on and how frequently the university was targeted.

The study then wanted to discover how many total recipients each subject category had. Each of the 240 emails was sent to its own unique recipient pool. Some emails were sent to only a few recipients while others were received by over 1000 of the university’s account users. It is important to see the distribution of recipients who received the determined subject grouping. Table 4 shows the frequency distribution recipients by which email topic, labeled in this study as “Subject Theme”, they received. In the study overall, a phishing emails was sent to a recipient to one of the 240 phishing emails 181,346 during the collection period. Due to not collecting data on the specific recipients, it was unsure how many users were attacked with multiple phishing emails. Each received phishing email was recorded, even if the same recipient received multiple emails throughout the process every incident was seen as independent of previous ones. The results showed that there were overwhelmingly more recipients who received “Webserver” emails than the other categories. 60,462 individuals within the university email system received an email of webserver type topics. From the “Recipient” results alone it appeared that more recipients were targeted with “Webserver” emails than any other. However, the results from the subject email count should be viewed alongside how many recipients received each type. “Webserver” was also the “Subject Theme” that had the highest number of incidences recorded. Even though “Webserver” appeared to have more recipients than the other subject categories, it should be noted that it also had more occurrences than the others. Together, the “Email Count” and “Recipient” columns of Table 4 revealed a very important trend in how the recipients are being attacked in the university email systems.

Table 4: Subject Email and Recipient Frequency

Subject	Email Count	Percent	Recipients	Recipient Percent
Financial	9	3.8	7,089	3.9
Helpdesk IT	61	25.4	25,133	13.9
Login Password	49	20.4	52,133	28.9
Miscellaneous	38	15.8	36,302	20.0
Webserver	83	34.5	60,461	33.3
Total	240	100	181,345	100

The dependent variable of the four emotion categories can also help provide useful information on how academia is targeted with phishing emails. Table 5 showed the distribution frequency of the four “Email Theme” categories the 240 emails classified into email count and the total number of recipients that received each email. Emails classified as targeting the emotion of curiosity from recipients were sent the most as seen with the 100 cases of “Curious”. The lowest emotion category that was recorded was “Greed”. In order to make sure a phishing email was seen, social engineers write the title/subject of the email in such a way that recipients feel compelled to open it when they see it. Academia users were typically using their account for work purposes so emails that appear to be official might be responded to urgently with a feeling of anxiousness or curiosity.

When looking at the “Recipient” column in Table 5, the exact recipients count total each of the four emotion classifications had over the collection period. The recipients were recorded on an individual incident level so each time one of the phishing emails was received was recorded as a recipient incident. “Greed” had only 5,090 recipient incidences recorded and had the lowest number of phishing emails sent. Table 5 also showed that phishing emails targeting “Curious” emotions had the most overall recipients with 106,313 total. A possible reason might be that social engineers feel more people would react to emails that sparked curiosity. However the high number of recipients can also be simply from the “Curious” emails having a larger over all count. “Anxious” emails had a much lower number. A possible conclusion for this result was that more spear phishing attempts were made to individuals who received emails targeting anxious feelings in recipients. Of all of the “Emotion Theme” categories, emails intending to create a feeling of anxiousness were sent to the most recipients within the sample. In spear phishing cases social engineers were after a very specific person’s information or type of knowledge so they attack those who can provide desired information.

Table 5: Emotion Email and Recipient Frequency

Emotion	Email Count	Email Percent	Recipients	Recipient Percent
Anxious	90	37.5	46,039	25.4
Curious	100	41.7	106,313	58.6
Confused	42	17.5	23,903	13.2
Greed	8	3.3	5,090	2.8
Total	240	100	181,345	100

Next the study focused on the recipient information. The recipient information referred to the position descriptions used by the university. It did not include information on individual people but instead on how different positions were targeted through the university. Table 6 showed the frequency distribution of the "Role" of each recipient in the university. Any recipients classified as "Employee" in the university email system were targeted the most with phishing emails. Employees are individuals who were most likely actively on campus on a regular basis and thus checking their university email accounts often, it was reasonable to understand that employees would be sent the most phishing attempts. "Affiliate" individuals received the least amount of phishing emails with 3,473 incidences reported. Affiliates are typically those individuals who do certain projects and task with a university but were not primarily involved on a daily basis.

When a social engineer was trying to get into a university system, the most logical victim would be an employee versus someone who is affiliated. An employee will have a greater chance of using their email account before affiliates. "Students" who were university employees were targeted heavily with 24,827 phishing emails received. Since the university does not employ a large numbers of students, social engineers might have wanted to target student employees because a student might not be as careful as employee when opening emails.

Table 6: Role of Recipients Frequency

Role	Frequency	Percent
Employee	135,893	74.9
Affiliate	3,473	1.9
Department	6,311	3.5
Retired	10,841	6.0
Student	24,827	13.7
Total	181,348	

Table 7, the frequency table of the "Title" variable, showed if recipients of the 240-phishing emails are considered a faculty member or not. Faculty members were those who teach or lecture the university students and make up an estimated 1/10 of the total number of paid individual of the sampled university. Given the low number of employed faculty members, Table 6's distribution of more recipients being "Non-Faculty" with 139,070 cases is understandable. However the "Faculty" recipients should not be overlooked. Faculty members are individuals who might be recipients of grant and funding money for research. Social engineers could have attacked faculty members for their grant money, salary, or to gain access to intellectual property created through research. However, if social engineers wanted to gain entry into the university system as a whole, getting access to any user's account information will do. In some cases, social engineers wanted to also target "Non-Faculty" in hopes that the person is not as educated and alert as a professor would be when receiving emails. Those who were faculty members might be more likely to get spear phishing emails because of their importance within the education of the university.

Table 7: Title of Recipients Frequency

Title	Frequency	Percent
Faculty	42,275	23.3
Non Faculty	139,070	76.7
Total	181,345	100.0

The “Division” frequency distribution in Table 10 shows the breakdown of the specific area within the university that each recipient of the phishing emails is employed under. The largest number of recipients for the “Division” variable, 88,548 recipients, belonged to the “College” division. “College” groups together professors, administration, and any other jobs that are classified as being a part of one of the university’s college degree programs. “Library” university users were targeted the least with phishing emails. There were considerably fewer university individuals that work in library type positions than in the colleges and administration. However, “Library” individuals might be targeted because of the access they have to documents that were not readily available to everyone. “Administration” individuals might be targeted by social engineers because of the unique access to the university system that they might have. Every one of the recipients in the five different “Division” categories was targeted with phishing emails for different reasons, which makes all of them important to the university security. Neglecting one of the job divisions could result in a complete system wide-compromise because that section was attacked and not properly secured. Table 8 showed that individuals are attacked by phishing attempts in every area of a university.

Table 8: Division of Recipients Frequency

Division	Frequency	Percent
College	88,548	48.8
Administration	47,039	25.9
Library	3,630	2.0
Affiliated	21,257	11.7
Extension	20,871	11.5
Total	181,345	100.0

In Table 9, a bivariate correlation test was used to determine if any significant relationships existed between the time variables and the five “Subject Email” categories. The correlation of the 2-tailed test was shown to be significant if the value was observed at the level of .05. The 2-tailed test observed no significant relationships between any of the “Subject Emails” categories and the time variables of “Duration” and “Day Sent”. The correlation 2-tailed value between the subject categories of “Financial”, “Miscellaneous” and “Webserver” and the “Month Sent” showed there was a significant relationship observed between them. The significance showed for the most part that the month phishing emails were sent has an impact on the subject category. Between “Month Sent” and “Financial”, “Helpdesk IT”, and “Miscellaneous” the Pearson Correlation was observed as a negative value. This negative value showed that the majority of the three categories had emails sent during the beginning of the collection period closer to June than in the months closer to December. The strong positive Pearson Correlation value for “Webserver” and “Login Password” showed the emails were sent mostly at the end of the collection period closer to December than in the earlier months.

Table 9: Correlations between Subject Emails and Time Variables

		Duration	Day Sent	Month Sent
Financial	Pearson Correlation	.014	.080	-.181**
	Sig. (2-tailed)	.832	.214	.005
	N	240	240	240
Helpdesk IT	Pearson Correlation	.108	-.042	-.046
	Sig. (2-tailed)	.095	.519	.476
	N	240	240	240
Login Password	Pearson Correlation	-.067	-.049	.102
	Sig. (2-tailed)	.303	.448	.115
	N	240	240	240
Miscellaneous	Pearson Correlation	.022	-.016	-.423**
	Sig. (2-tailed)	.732	.809	.000
	N	240	240	240
Webserver	Pearson Correlation	-.065	.060	.353**
	Sig. (2-tailed)	.317	.355	.000
	N	240	240	240

The "Emotion Emails" were then run in a bivariate correlation test with the three time variables in Table 10. Across all of the emotion groups, the time variables of "Duration" and Day Sent" were not shown to have a significant relationship. However, much as was seen in Table 9, the "Month Sent" time variable was reported as having significance with most of the emotion categories. The Pearson Correlation for the emotion "Anxious" and "Month Sent" were negative which means that more "Anxious" emails were mostly sent to recipients during the beginning of the collection period versus the end. Therefore, phishing emails appealing toward anxious emotions were sent closer to June than to December. The other three emotion categories of "Curious", "Confused", and "Greed" reported a positive correlation. The positive correlation showed that emails related to these three emotions were sent more toward the end of the collection period.

Table 10: Correlations between Emotion Emails and Time Variables

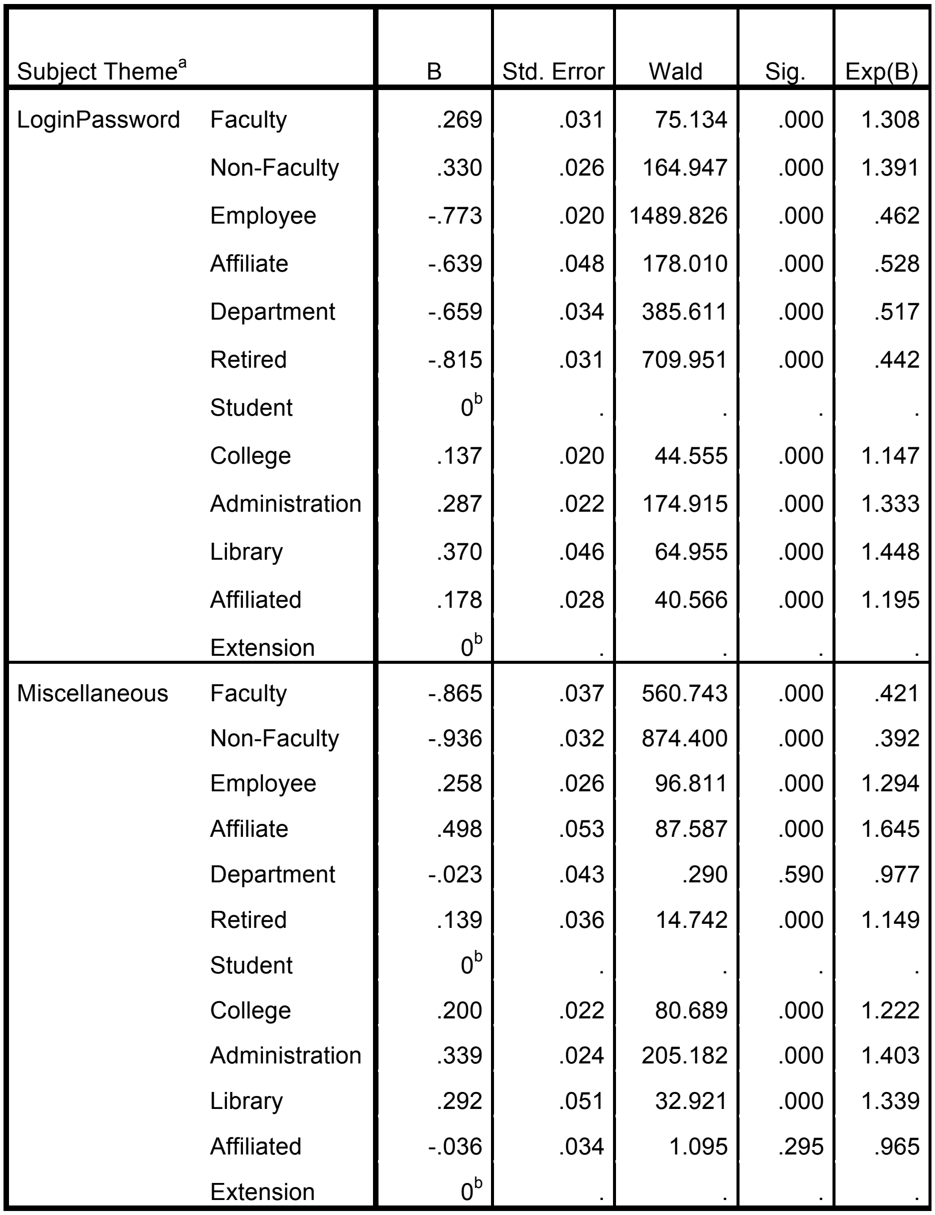
		Duration	Day Sent	Month Sent
Anxious	Pearson Correlation	.018	-.023	-.242**
	Sig. (2-tailed)	.778	.723	.000
	N	240	240	240
Curious	Pearson Correlation	-.027	.036	.033
	Sig. (2-tailed)	.677	.583	.611
	N	240	240	240
Confused	Pearson Correlation	.050	-.047	.140*
	Sig. (2-tailed)	.441	.473	.030
	N	240	240	240
Greed	Pearson Correlation	-.081	.063	.265**
	Sig. (2-tailed)	.212	.334	.000
	N	240	240	240

To determine if the dependent variables and independent recipient variables have a relationship with each other, multinomial logistic regression tests were used. The analysis of the "Subject Theme" categories possible relationships with the recipient information of "Title", "Role", and "Division" was found in Table 11. "Webserver" subject category was distinguished as the reference category because it was the subject category that had the most incidents occur, as seen in Table 4. With the regression for "Financial" relative to the reference group, most of the independent variables were significant given the others are in the model except the variables of "College", "Administration", and "Library". The "Affiliate" variable for "Financial has an odds ratio of 11.073. Therefore, recipients who were classified as being an affiliate were 11x more likely to receive financial emails.

For the "Subject Theme" multinomial logistic regression, the "Webserver" category was used as the reference category. The regression for "Helpdesk IT" showed that only one of the independent a variables "Library" was not significant while the other variables did have a significant relationship. The "Affiliate" variable had the largest odds ratio, showing affiliates recipients was over 4x more likely to receive helpdesk emails than the other groups. The regression for "Login Password" showed that all of the independent variables were significant but none had a huge percentage of occurrence compared to the others. The regression for "Miscellaneous" showed that the variables of "Department" and "Affiliated" were not significant with the miscellaneous emails. Across all of the five subject categories, the "Title" and the "Role" the recipient held within the university appeared to have a statistical relationship with why they received those particular emails apart from the "Department" cases within "Miscellaneous" emails. The "Division" however appeared to have less of a significant relationship with the subject categories. The lack of significance could imply that social engineers looked at an individual's specific role and title within the university to know what type of email will be more successful.

Table 11: Parameter Estimates for Subject Theme Multinomial Regression

Subject Theme[a]		B	Std. Error	Wald	Sig.	Exp(B)
Financial	Faculty	-3.795	.100	1451.966	.000	.022
	Non-Faculty	-3.783	.094	1626.446	.000	.023
	Employee	1.785	.089	405.622	.000	5.959
	Affiliate	2.405	.113	454.761	.000	11.073
	Department	1.532	.108	199.475	.000	4.627
	Retired	1.202	.106	129.257	.000	3.326
	Student	0[b]	.	.	.	.
	College	.026	.039	.451	.502	1.027
	Administration	.021	.042	.251	.616	1.021
	Library	-.180	.103	3.069	.080	.835
	Affiliated	-.319	.073	19.050	.000	.727
	Extension	0[b]	.	.	.	.
Helpdesk IT	Faculty	-1.991	.047	1800.475	.000	.137
	Non-Faculty	-1.860	.042	1952.200	.000	.156
	Employee	1.033	.038	756.566	.000	2.808
	Affiliate	1.501	.062	579.607	.000	4.485
	Department	.510	.057	80.134	.000	1.665
	Retired	.870	.047	344.632	.000	2.386
	Student	0[b]	.	.	.	.
	College	.104	.024	18.448	.000	1.109
	Administration	.209	.026	66.474	.000	1.233
	Library	.065	.058	1.246	.264	1.067
	Affiliated	-.379	.043	78.563	.000	.684
	Extension	0[b]	.	.	.	.



Subject Theme[a]		B	Std. Error	Wald	Sig.	Exp(B)
LoginPassword	Faculty	.269	.031	75.134	.000	1.308
	Non-Faculty	.330	.026	164.947	.000	1.391
	Employee	-.773	.020	1489.826	.000	.462
	Affiliate	-.639	.048	178.010	.000	.528
	Department	-.659	.034	385.611	.000	.517
	Retired	-.815	.031	709.951	.000	.442
	Student	0[b]	.	.	.	.
	College	.137	.020	44.555	.000	1.147
	Administration	.287	.022	174.915	.000	1.333
	Library	.370	.046	64.955	.000	1.448
	Affiliated	.178	.028	40.566	.000	1.195
	Extension	0[b]	.	.	.	.
Miscellaneous	Faculty	-.865	.037	560.743	.000	.421
	Non-Faculty	-.936	.032	874.400	.000	.392
	Employee	.258	.026	96.811	.000	1.294
	Affiliate	.498	.053	87.587	.000	1.645
	Department	-.023	.043	.290	.590	.977
	Retired	.139	.036	14.742	.000	1.149
	Student	0[b]	.	.	.	.
	College	.200	.022	80.689	.000	1.222
	Administration	.339	.024	205.182	.000	1.403
	Library	.292	.051	32.921	.000	1.339
	Affiliated	-.036	.034	1.095	.295	.965
	Extension	0[b]	.	.	.	.

The multinomial regression for the “Emotion Theme” dependent variable in comparison to the independent variables also showed interesting results. The emotion category “Curious” was distinguished as the reference category for the regression test because it received the most emails of all of the four emotion categories. The reference category for the regression test was determined by the frequency values in Table 5. “Curious” subject category was distinguished as the reference category because it was the subject category that had the most incidents occur in the data set. With the regression for “Anxious”, the “Division” categories of “College”, “Administration”, and “Library” were not significant while all other variables were. The regression results showed that all of the independent variables were significant for the “Confused” and “Greed” emotions. Across all of the emotion categories, the “Title” and “Role” independent variables were significant. “Title” and “Role” of the recipients was more significant than the “Division” to social engineers when they selected targets. When looking at just emails that appeal to confusion, the recipients labeled as “Employee” and “Retired” had larger percentages of receiving “Confused” emails than the other university users. When looking at recipients who received “Greed” emails, “Employee”, “Department” and Retired” individuals where likely to receive more than the other recipient categories. Recipients who were classified as having a “Department” role were 28x more likely to receive phishing emails that appealed to greed. Across all of the emotion categories, recipients who were labeled as “Employees” were likely to receive the most phishing email.

Table 12: Parameter Estimates for Emotion Theme Multinomial Regression

Emotion Theme[a]		B	Std. Error	Wald	Sig.	Exp(B)
Anxious	Faculty	-2.254	.035	4174.300	.000	.105
	Non-Faculty	-2.257	.031	5228.084	.000	.105
	Employee	1.637	.027	3567.256	.000	5.142
	Affiliate	1.903	.046	1727.829	.000	6.704
	Department	1.390	.039	1241.709	.000	4.015
	Retired	1.403	.035	1590.586	.000	4.068
	Student	0[b]	.	.	.	.
	College	.026	.019	1.859	.173	1.026
	Administration	.006	.020	.081	.776	1.006
	Library	-.007	.043	.027	.868	.993
	Affiliated	-.447	.032	193.190	.000	.640
	Extension	0[b]	.	.	.	.
Confused	Faculty	-3.353	.053	4004.463	.000	.035
	Non-Faculty	-3.540	.050	5111.960	.000	.029
	Employee	2.457	.046	2804.840	.000	11.666
	Affiliate	1.761	.071	620.540	.000	5.820
	Department	1.766	.060	856.435	.000	5.848
	Retired	2.486	.053	2226.109	.000	12.008
	Student	0[b]	.	.	.	.
	College	-.185	.022	68.936	.000	.831
	Administration	-.470	.025	361.747	.000	.625
	Library	-.365	.056	41.848	.000	.695
	Affiliated	.097	.039	6.284	.012	1.102
	Extension	0[b]	.	.	.	.

Emotion Theme[a]		B	Std. Error	Wald	Sig.	Exp(B)
Greed	Faculty	-5.949	.135	1931.778	.000	.003
	Non-Faculty	-5.982	.130	2127.248	.000	.003
	Employee	2.803	.121	532.340	.000	16.495
	Affiliate	2.026	.185	120.094	.000	7.584
	Department	3.350	.128	688.733	.000	28.516
	Retired	2.500	.135	341.825	.000	12.188
	Student	0[b]	.	.	.	.
	College	.325	.054	36.550	.000	1.384
	Administration	.503	.055	84.176	.000	1.654
	Library	.698	.097	52.067	.000	2.010
	Affiliated	.204	.085	5.844	.016	1.227
	Extension	0[b]	.	.	.	.

CHAPTER 5 DISCUSSION

5.1 Research Findings Importance

A phishing email's purpose is to convince the victim to open it and provided confidential credential or information. After comparing the 240 emails, it became apparent that phishing emails follow a basic format of composition that can provide key indicators for recipients. The content theme analysis measures of this study provide practical procedures that can be applied by an individual to avoid falling victim to future phishing emails. First, a user should be cautious of any email that is not personally addressed and has a broad signature. Second, users should also be aware of spelling and grammatical errors that appear in any email. Grammar mistakes are a major indicator of a phishing email. Thirdly, if an email says to follow a link to enter an account, the user should never use the link provided. Even if the email is asking the recipient for a request to check something on the link, the recipient does not have to comply. The recipient may feel the need to respond but should think critically about the request before doing so. Instead, the user should open a new tab in their browser to enter their account. The user may have a legitimate account they hold mentioned in the email but the link may take them to a scraped page to steal their information. If the user is concerned about the status of their account, they can easily check it by logging in separately from the email.

The first important part of an email’s presentation is the title. Being that it is the first part of the phishing email seen by a recipient, the subject/title line of each email is important to study. The majority of phishing emails in the sample had capitalized subject lines and correct spelling. However, a few of the phishing emails displayed titles that are not capitalized and contained minor spelling errors. A misspelled subject/title should be an instinct indicator to recipients that the email is a phish. However, if the recipient still opens the email to read it, there are more indicators that they should be alert for when looking at the email body. The phishing emails typically have an attention catching subject/title that causes an emotional response in the target, causing them to feel compelled to open the email. The data sample showed that language in the body of the email could also indicate to recipients that it is a phish. Grammatical structure and spelling errors are found in many of the phishing emails from the sample. Many of the emails had small grammar issues, such as not capitalizing the pronoun “I” or capitalizing other words that make the sentence grammatically incorrect. Some of the phishing emails in the sample displayed sentences that, while they are technically correct, are not written in a native English dialect. Illogical sentence structure can happen when the social engineers who create the email are from a foreign country and use a rough translation to create the email. Odd sentence structure should be another immediate indicator for a recipient that the email may be fraudulent.

The results of this study show that spelling errors are very important to detecting phishing emails. Spelling errors can be an easy indicator for recipients that the email is fraudulent. With only 28 of the 240 gathered emails containing spelling errors, this shows a level of sophistication among social engineers. Social engineers are aware that spelling errors are easily detected. By making phishing emails grammatically correct, social engineers are increasing the chances that recipients believe the emails to be legitimate.

The 240 emails are also divided into regular phishing emails and incidents of spear phishing. Spear-phishing emails are sent to a specifically selected number of individuals. Most of the phishing emails in this sample are sent to a large number of people based broadly on the recipient's job description in hopes of getting a response. Spear phishing emails are more selective and deliberate. Spear phishing emails are used when social engineers want to get information from identified targets. For the study, any spear phishing email was categorized as having less than or equal to 50 recipients. Of the 240 emails in the sample, 98 are classified as spear phishing attempts. When less than 50 individuals are sent a spear phishing email, it may means social engineers select them specially in hopes of getting specific information from them. Spear phishing emails are very important to study because they help IT departments understand why certain individuals are attacked. Spear phishing is not a main focus of the present study but should be expanded upon in future studies focusing on phishing email data.

Social engineers are attacking academic institutions with phishing emails every day for a variety of reasons. The way social engineers get the information they need is through requesting favors/tasks from people they contact. With the technology and Internet growth, sending emails has become a common way for social engineers to ask victims for information. As seen with the Social Exchange Theory, a recipient is more likely to comply with what the email says when they feel the sender has already helped them in some way. Social engineers try to get recipients into obligatory relationships in phishing emails where recipients feel they have to respond using social exchange principles. One reason could be that attackers are trying to gain access to a university user's personal credentials in order to steal payment information. Another possible reason academia is targeted is because they create intellectual property that is coveted by other countries and competitors.

The time variables provide important information on how academia account users are targeted by phishing emails. The days that phishing emails are first sent to recipients shows when social engineers will most likely choose to attack their victims. In Table 1, most of the phishing emails are sent during the workweek business days. Although 21 of the 240-phishing emails collected are sent during the weekend, the vast majority are sent during the workweek. Social engineers might choose to send emails Monday through Friday because that is the day when users will be checking for the upcoming workweek. However, Table 1 also shows that while the workdays overall had significantly more incidences recorded, Tuesday and Thursday had less than the other three workdays. An exact explanation for this phenomenon is unknown, but could be a factor of the social engineers schedules. The weekend days have very low numbers of incidences, which shows that social engineers are aware that sending a phishing email on a weekend day has a low likelihood of being opened. Social engineers want to maximize the chance the email is opened before IT departments stop it, therefore sending emails on weekdays is more efficient. The university IT department can take these results and use them when preparing education training as well as improving spam filters.

The particular month that a phishing email is sent to recipients offers valuable insight into when university account users are targeted. The frequency distribution in Table 2 shows that the middle of the collection period had a relatively steady inflow of phishing emails. From July to October over 40 emails were sent each month. This distinct time period is during the bulk of the Fall academic semester, when the phishing emails will have the greatest chance of being seen quickly by the recipient. A social engineer knows that a college campus has the most people actively using their emails when they are during the Fall and Spring semesters. The number of incidences began declining in November and is the lowest in December. November, while still in the academic semester, is the beginning of the holiday season as well as being toward the end of the semester. For the majority of the month of December, the university is typically on holiday and closed for business. During this time, employees will be accessing their email accounts less. Therefore, sending phishing emails in December reduces the chances of recipients opening them before IT officials blocks them. In doing their background research before sending their attacks, social engineers will use information on holidays and semester workdays to make sure they target individuals when they will be using their email accounts the most.

The “Minute Duration” distribution results in Table 3 show the total amount of minutes that all of the phishing email cases for each subject category remained in the university system. While the differences may appear very significant across the subjects, it should be noted that the case number for each subject was not equal. Only nine “Financial” phishing emails are in the sample while there are 83“Webserver”. This large difference in the cases can be a factor in the low count of total minutes of duration. Combining Table 3 with Table 4 can offer a more holistic view of how long certain phishing email go unnoticed by IT spam filters. However, Table 3 does still show that phishing emails do on average go unnoticed by IT departments for a considerable length of time. Every minute a phishing email is active is another chance for social engineers to compromise the university system. Unfortunately the IT department only learns of an active phishing email in the system if recipients report it. An important part of stopping phishing emails is having email account users report phishing attempts in a timely fashion. University email users need to take the responsibility of being proactive regarding phishing emails. If the email is a phishing email, the IT department officials will be able to block the email before any more are sent to other university email users. When the recipients are proactive, IT departments can also be proactive instead of reactive.

The “Subject Email” variable provides information that distinguishes incidences of content focused emails versus others. Knowing which emails go unnoticed the longest can help IT edit their current spam filters and adapt them to be more effective. Table 4 shows the frequency count of individual recipients for the 5 “Subject Theme” categories as well as the email count for each category. The results show that there were more recipients of the phishing emails related to “Webserver” subject matter than the other subject theme categories. The “Helpdesk IT” and “Login Password” categories also are showen to have a high number of occurrences. “Emails containing content related to IT, email, and passwords are ones that university users might receive legitimately from their university administration and IT departments. When constructing a phishing email that will have the most success, social engineers want to create one that has content to attract victims to believe it but to also seem legitimate enough to evade spam filters. Sending a university employee an email talking about a financial matter will be flagged as fraudulent faster than one that asks for a password reset because the university does not deal with financial information often with its users (Identify and report Scams and Frauds, n.d.). Given the distribution of incidences, the university IT department can use the information in their education programs to inform account users of the different forms that phishing emails take.

The recipient frequency column in Table 4 shows that there were a total of 181,345 recipients of phishing emails during the collection period. Given the number of non-students associated with the university is typically estimated around 10,000 every year, the high number of recipients means that a single individual could have received multiple phishing attempts during the collection period. Every time a recipient received a phishing email it was recorded as an independent incident. Due to privacy of the user’s personal information, this study does not track the numbers each individual university account holder used so it is unsure if certain individual recipients are targeted more than others. The study also is unsure how many social engineers are involved in sending the 240 emails, so it is unsure if a single social engineer kept attacking the same individuals. However, 181,346 recipients of phishing attacks can be seen as too many phishing attacks allowed into the system. The more phishing emails that a person receives, the less likely they will be able to tell that each is fake and they might believe one appears more realistic than the others. Social engineers could be attacking the same people with multiple different phishing attempts. The distribution of the total number of recipients for each of the "Subject Themes" helps show what type of emails are sent to the most recipients but the data should be compared proportionally with the total subject "Email Count" since the number of incidences recorded for each of the subjects is not equal.

Phishing emails are sent to the university recipients by social engineers in the hopes of gaining access to confidential information. However, once social engineers send the phishing email they can only wait in anticipation to see if the email is successful. To ensure success, social engineers must be certain that the email entices the recipient enough to open and respond to it. Social engineers rely on the human behaviors seen in the Social Exchange Theory to entice recipients of phishing emails to respond. SET explains the principle that individuals will be likely to want to reciprocate any favor they receive. Generally, humans will behave logically and want to help themselves. If an individual is offered a reward in exchange for a small cost, they are likely to oblige to the exchange. Social engineers can write the phishing emails to appear as if they have already helped the recipient in some way, which makes the recipient feel indebted. Once the recipient feels as if they owe the sender, they will want to provide what they can in return. A simple way to make victims feel compelled to open a phishing email is to make the recipient feel a certain emotion trigger when reading it. Emotions have a huge impact on human behavior and can be used to manipulate a person more successfully.

The 240 emails are grouped into emotions that the researcher felt are the most prevalent when reading the emails. Table 7 focuses on seeing if any of the four emotions are more frequently invoked than others. The results show that academia personnel are typically sent emails that appeal to their curiosity and feeling of anxiety more than feeling of greed or confusion. Recipients might be more inclined to take action on am email that has aroused their feeling of anxiety because they feel the need to urgently resolve the situation. When a person is made anxious, they may lose some rational and logical ability to assess the email for legitimacy because they want to urgently address the matter discussed in the email. Emails targeting anxious and curiosity emotions match better with typical office-type emails an individual may receive. This can also aid in explaining why academia may receive less emails appealing to emotion of greed. Phishing emails appealing to greed will be very financially heavy and users might not be expecting financial emails sent to their work accounts. Instead, if a user receives an email talking about money or bank account while on their university email account, they may be more proactive in realizing it is fraudulent.

Much like with the subject categories recipient distribution, the “Emotion Theme” recipient distribution results show a similar pattern to the groups into which the 240 emails are classified. The results of the “Recipient” column for Table 5 show that more recipients received phishing emails that tried to invoke a feeling of curiosity from recipients more so than the other groups. However, the data results from this table should be analyzed alongside the results from the “Email Count” column for an accurate depiction of the ratio between emails sent and the recipients. The emails appealing to anxious feelings are sent almost as often as those appealing to curiosity but have fewer recipients. A possible explanation is that a feeling of anxiety was more likely to be targeted in incidences of spear phishing. If a social engineer wants to be sure to get certain information form a particular person as quickly as possible, creating a feeling of anxiety might have more effect than other emotions. Phishing emails appealing to the recipient’s greed are sent the least to the university users.

The “Role”, “Title”, and “Division” that each recipient is labeled as in the university system offered valuable insight into how social engineers pick targets in academia. University email account users who tend to be classified as “Employee” are targeted the most with phishing emails. If social engineers know nothing about the inner workings of the university, they can look around on the university’s website they to find individuals listed as employees. Given that employees typically have access in the system, a social engineer might broadly pick a list of employees to target. All social engineers need is one person to comply to gain access to the system. Social engineers also will be more likely to choose an employee over someone who is retired and no longer on the system. Those who fall under the category of “Non Faculty” are frequently targeted with phishing emails than faculty members. The large difference between these groups can be greatly impacted by the fact that the university employs a far larger amount of “Non Faculty” individuals than “Faculty”. The faculty group is a much small group of individuals. Only those individuals with professor or lecturer in their university title are included in the “Faculty” group. The “Non Faculty” group covers a wider range of position titles within the university such as administration, janitor, assistant, etc. The data shows that there are a larger number of faculty members being targeted, which IT security officials do need to focus on. Faculty members are more likely to make intellectual property that is valuable as well as have access to large grant money accounts, which make them attractive targets for social engineers.

The "Division" of the recipients of phishing emails perhaps provides the most important information in creating security policies for academia. The distribution of the "Division" variable lets security officials know who is being targeted with phishing emails within the university. In relation to the university as a whole, the "Division" shows the breakdown of recipients attacked on a global level. The five categories show the distribution of how major university job positions are attacked. Given the results in Table 8, IT security officials can learn that they should make it a priority to protect their users against phishing emails. Making sure that security policies and education reach college personnel is important, however the other groups cannot be overlooked. Every phishing email received by a university account holder is a chance to have the system compromised, regardless of who they are. The distribution of which "Division" has the most individuals attacked may provide information on how social engineers selectively target individuals for certain information through spear-phishing. Overall the distribution shows that every area of the university is a potential target.

When the time variables are compared with the dependent "Subject Emails" variable, the Pearson Correlation shows that phishing emails with different subject content are sent at different times of the academic semester. The correlation test in Table 9 shows that the months sent are the only time variable that is related to the subject content of the phishing emails. As

Conheady mentions in *Social Engineering in IT Security: Tools, Tactics, and Techniques*, the data shows the month a phishing email is related to the subject content. An explanation for this could be that social engineers send phishing emails with certain subject lines during months that the email might be better received. For example, sending emails with IT related topics might have more success at the beginning of the school year because users want to make sure that they comply with any system updates for the school year. In particular, the subject emails related to “Financial”, “Helpdesk IT”, and “Miscellaneous” matters appear to be sent toward the beginning of the Fall semester. The other subject categories of “Login Password” and “Webserver” are sent mostly during the end of the Fall semester. The patterns of what months certain subject categories are being sent can be used when making educational training for university users. IT officials can use the correlation results when creating security training and help inform potential targets when certain content oriented phishing emails might be more prevalent.

As with the “Subject Emails” variable, the bivariate correlation results in Table 10 shows that the “Month Sent” is shown to be significant with the “Emotion Email”. The other variable comparisons did not have significant relationship findings from the significant value (2-tailed). Also, the Pearson Correlation results did not show a strong relationship between any of the variables. This shows that social engineers do consider what month it is when they are designing an email to send to individuals. Some phishing emails have more success during certain months, such as when employees access their accounts. In fact, “Anxious” emails show to be sent more frequently at the beginning of the Fall semester than the end. The other emotion categories are sent mostly toward the end of the Fall semester. A conclusion can be drawn from the correlation tests that social engineers do factor in what month it is when sending specific subjects and emotion rich emails but not the day of the week. Table 9 and Table 10 both show that the three different time variables are not significantly related to each other. This means that social engineers who create the phishing emails did not follow a distinct pattern of the day the specific content-based email is sent, except for the month. When IT officials are improving current spam filters and policies, they need to make sure they focus on specific months being high risk for certain subject and emotionally charged content.

Time factors, other than the month sent, do not appear to be as significant to the content of the email. Even though the content does not appear to impact the time of day a social engineer sends certain emails, the month does show a potential relationship. The month signifies when university users are most likely to use their account while the exact day of the week is simply a further break down of that time frame. Social engineers appear to consider the month they are sending an email with specifically topic based content but do not consider the day of the week. However, the day of the week is still important to when the total 240 phishing emails enter the system as one group. Social engineers do not focus financial based emails on only one day of the week specifically, but tend to send any of their emails broadly during workdays. Social engineers goal is to get the victim to receive the email when they are most likely to be at work.

The multinomial logistic regression analysis test shows the predictive analytics that a dependent variable has a relationship with one or more categorical independent variables. In Table 11, the "Subject Theme" into which the 240 emails are grouped are compared with the independent variables describing how the recipients are viewed within the university. The results show that in most cases social engineers will pick recipients to target with certain phishing email content based on the positions they hold within the university. The "Title" of the recipient was shown to be significant across all of the subjects. A significance with the "Title" shows that social engineers would consider if the intended victim is a faculty member or not when they sent out phishing attacks. Recipients who are categorized as "Affiliate" are 11x more likely to receive financial related emails. Affiliates of the university should therefore be made aware of their potential for possible financial phishing email attacks. The "Role" variables are also significant across the subjects except for "Miscellaneous" with the "Department" variable. Again, social engineers might specially pick victims to target with phishing based on the role they have within the university. Of all of the independent variables, the "Division" variable have the most insignificant results when compared with the subjects. "Login Password" is the only category that showed significance in all of the "Division" categories. These results show that the extent to which the employee belonged is less of a factor regarding their receipt of certain subject-themed phishing emails.

The multinomial regression for the "Emotion Theme" dependent variable in comparison to the independent variables also shows how social engineers pick their victims for certain emails. Overall, the results show most of the variables are consist with the emotion groups, which can be a factor of the volume of recipients. The "Title" and "Role" variables are significant for all of the emotion categories. The "Division" variables are significant in most of the emotions overall except "Anxious". This shows that the "Title" and "Role" an intended victim has is more significant than the "Division" to social engineers when selecting who to target with emotionally charged phishing emails. In particular, individuals who are classified as “Employee”, “Department”, and “Retired” showed large percentages of getting phishing emails that appeal to the recipient’s greed. Both of the regression tests show that whether the email has a certain emotion pull or a subject topic displayed, social engineers do select the recipients based on the position that they hold within the university. IT departments can use this information to prepare security training particularly for university recipients who have the highest risk of being sent phishing emails.

5.2 Predicting and Preventing Remote Social Engineering

Companies and industries can purchase top-of-the-line IT security software to protect against cyber-attacks but it will not protect against the human component. As technology progresses, criminals are using social engineering tactics aimed at the employees in order to gain access into the individuals’ and the industries’ personal information. Based on their knowledge of human interaction principles seen in Social Exchange Theory, a social engineer simply needs to know how to word their request in the right manner in order obtain a high rate of compliance. In most cases, the recipient does not even know that they are being taken advantage of until it is too late. There is never going to be a completely effective policy or education program to protect against all remote social engineering, but having a basic knowledge of certain signs of attacks and prevention measures can help reduce the success of social engineers. Employers and organizations need to equip their employees with knowledge to be more aware of possible attacks. Simply being aware of what to watch out for and following some easy tips can drastically improve a company’s chance of not being breeched by social engineers.

All universities and academic institutions should make protecting their intellectual property a top priority. Many professors and researchers that are a part of the university create intellectual property from the findings of their studies. Many of these studies are also funded privately and through government agencies, which make the findings very important to protect. By stealing intellectual property, social engineers are stealing research findings, ideas, and inventions that can be copyrighted. Social engineers might be hired by private parties in other countries in order to learn new advances in research areas that United States has discovered, which can bring into concern national security for government funded projects. Professors and university employees alike will continue to receive fraudulent emails and fall victim unless they are educated with knowledge of phishing emails.

To gain access to the university system, all social engineers need to do is to gain the credentials from a single system user. This user can be anyone from a college dean to a custodial worker, which makes cyber security important to every user on the network. However, many computer users do not have any security education to help them avoid becoming a victim. Through proper education and training, users can become equipped with adequate knowledge to be able to spot a phishing email attempt before they click on it. The statistical tests run in this study help to highlight areas that security training needs to focus on to help protect academic institutions from remote social engineering attacks.

In fighting against social engineering attacks, an organization needs to be aware of certain indicators of an attack, update security polices, provide employee awareness and education, and create plans regarding responses to an attack (Conheady, 2014). Predicting a remote engineering attack is difficult. Luckily, organizations can create defenses against them by educating their employees on indicators of attacks. The employees are the human element of the company and thus the weakest link in the security system. Recipients need to determine if the email is legitimate by critically analyzing the overall message presented in an email. There are easy indicators that organizations need to look out for to know if they are experiencing social engineering attacks. The first indicator is the person's attitude. A social engineer's attitude and presentation might reveal that something is amiss. If the person appears to be making an exaggerated effort to convince authority figures of legitimacy, it could be a strong indicator that they are trying to infiltrate the organization. Also, another indicator is if the person is particularly emotional sounds unnatural and overly social. Employees should be diligent and question anything that even slightly arises suspicion. A third indicator that employees need to watch out for is if someone they do not know is blatantly trying to establish a connection them. This connection can be established through use of personal information about the victim, over use of company terms, and even name-dropping to sound as if they are part of the organization.

Employees need to be able to question the nature of the request that someone unfamiliar is asking of them. One must consider if the person is authorized to access the information they want and if the request came in a natural way. If the employee is not used to having similar request then they need to be on alert. When in doubt, an employee can always contact their IT department. Also it is important to be alert if the request came at an unnatural business time such as after business hours or when the supervisor is out of the office. Social engineers might be pushy or urgent in trying to get the victim to complete an action. If they feel pressure that they might not be successful, social engineers might switch their actions to appear more authoritative or threaten negative consequences.

The data sample in this study provided some useful information in helping protect academic and other institutions against phishing emails. Many phishing emails are sent with similar themes and topics. Common phishes sent to academic institutions target recipients by phishing email content pertaining to academic matters such a webservers and IT alerts. Recipients of emails in academia need to be aware of the main subject areas that are sent most often to academic institutions. Phishing emails are seen to be sent out in mass numbers are impersonally addressed, might have bad grammar or spelling errors, and have strange wording.

Spear phishing attacks are a bit trickier to spot since they are specially designed to be convincing to particular people. In order to convince an individual to open an email, social engineers might research the person and construct the email in a relevant way. Individuals in the educational community, for example, need to be diligent to notice unfamiliar emails that end in ".edu". Many Social Engineers will use ".edu" email addresses to appear to be from a fellow scholar. Internet users should always check the email address to make sure it is correct or one they are expecting.

Every organization should have security policies and procedures in place. These policies should also be routinely assessed and updated to include current threats to the ever-evolving technical world. Protection procedures need to be written down and easily accessible to all personnel. If a procedure is heavy with technical terms, many of the system users will not be able to understand it. Procedures need to be easy to understand and read so the staff does not simply skip over them. Every industry and organization will have varying needs on finer operational details that are unique to them, however every security policy should at least cover a few major areas. There needs to be a clear outline of how employees need to handle waste such as old documents and computer storage devices. Especially if there are any secure documents stored on system users' computers, a certain procedure needs to be followed to dispose of the documents in a proper way. Whether it be shredding the documents or sending to a burner, an organization needs to dispose of documents that may carry confidential information appropriately to eliminate the possibility of dumpster diving. Dumpster diving is an easy way that social engineers can access details and otherwise useless information to a company but can be used to persuade their way into the building.

The most effective way to help protect technology users from attacks is to increase their awareness and education on phishing. Education programs need to be designed with the users in mind. Education about social engineering needs to be presented in a creative way that causes people to be interested in the material. If the education provided is too long and repetitive, it decreases the chance that the employees will pay attention and learn the information. The trick to creating a security program that will be the most effective is to think of who the targets are. People learn and react to things differently so creating a blanket program to be effective for everyone is impossible. However, there are a couple tips to consider when creating security policies that can help policies reach the most people positively.

Academic institutions need to prepare and deliver education on phishing emails and cyber vulnerabilities regularly to all of the users on the networks. One way that academic institutions can increase users' awareness of cyber attacks is to require regular training for everyone, including high-level administrators. If everyone is required to complete training, it will be valued and received in a better light. Second, IT personnel should consider requiring a mandatory password reset every few months. With a required password reset, any user accounts that could be compromised can be dealt with in a timely fashion. Third, a university can require that every individual who uses an email account on the university servers only use the account for school/professional use. The more casual sites (ie. Shopping, blogs, etc.) a user's email signs up for opens that account to more phishing emails simply because their email is more readily out on the internet. The less places that the email is used on the internet, the less likely it will be picked up by social engineers, given to other social engineers, and used to send remote attacks.

Companies and institutions should also consider hiring a social engineer to find security flaws and offer suggestions for improvement. A social engineer can spend a few days in the system and learn the flaws, which open them up for attacks. Alternatively, a social engineer could try to infiltrate the company and, when successful let management know every step of their process. This stimulated attack can give management a concrete and realistic view of the dangers of being breached. Another way a hired social engineer can test the system is to construct a phish using software tools and send it out to the company. If the recipients click on the link provided, the link will take them to a phishing awareness page. This form of education can be effective for some people because it causes them to learn from the mistake they made, but this form of education might anger others and cause them to feel betrayed or tricked.

Phishing emails are easily created and distributed by social engineers but they are also easy to avoid. If the recipient receives a phishing email and is unsure if it is legitimate, they can forward it to their IT department for confirmation. Many phishing emails contain hyperlinks that, when clicked, instantly downloads viruses to the user's computer or send them to a fake site to extract PII (personal identifiable information). If an email claims that an account the user holds is compromised, the user can simply open an Internet tab and login to the account regularly without clicking on any provided links in the email. Also, a user should never open an attachment in an email unless they know the sender and are confident it is safe. Many attachments carry malware and viruses that can infect the computer system if opened. If users remember to never open provided links and attachments (unless expecting the email), then the risk of compromising their account credentials is drastically lowered.

Phishing emails, although not done in person, are social exchange interactions between at least two separate parties. Human behavior has been found to be unfortunately predictable across the masses in regards to how individuals respond in social exchanges. For years, people have been able to take advantage of individuals in person, but technology has brought about the new ways of fraud. The phenomena of getting recipients to fall victim to phishing emails can be explained by the principles seen in the Social Exchange Theory. Social engineers can now construct a request to the recipient that can offer nameless rewards in place of cost that the recipient provides. In a phishing email, social engineers can also claim to have already helped the recipient in some way, which means that the recipient will feel obligated to reciprocate. An individual who receives a phishing email may see any request as a reasonable action and not of any high cost to them. Social engineers might not have any knowledge of the Social Exchange Theory but they use the principles and assumptions found in SET in order to attract victims to comply with their request. Phishing emails can be very detrimental to individuals and institutions alike, but with understanding of SET principles people can become more aware of any potential attacks.

CHAPTER 6 CONCLUSION

6.1 Looking Ahead

The Internet and the ways that users are attacked through social engineering means are ever-evolving. This study helps provide information that IT departments of organizations can use to focus their protection plans, however, continued research and education is always needed in the ever evolving technological age in which we live. As technology and security evolves, the social engineering techniques and methods evolve as well. Academic institutions can reduce their risks of being compromised through social engineering attacks, but they can never be completely secure. Academic institutions can effectively reduce their risk for social engineering attacks by increasing user education and implementing a few easy-to-follow security policy tips. The time spent on producing the education material and policy procedures will be very beneficial for the future. When in doubt, recipients should contact IT departments to make sure an email is safe to open. It is better to check multiple emails than to have one phishing email compromise the system because the recipient does not put forth basic security effort. By being a system user, each recipient is responsible for helping keep universities secure. As long as technology usage in academia continues to increase, academic institutions will be susceptible to phishing emails and other types of social engineering attacks.

6.2 Limitations

The results presented in this study are constrained by limitations in the collection process. The data sample of phishing emails collected is limited to those that the system recipients reported to the IT department of the university. If a phishing email is sent to a user but was never reported to the IT department, then the email was not collected for this study. Therefore, it is unclear what caused the reported emails to get noticed while others might not have been noticed. The study is also limited to the only system emails users collected are university employees. Student recipients, who are the majority of the university system users, are not collected unless they hold a campus job. Student account holders could have received the same phishing emails but due to Splunk only providing employee information, it is unclear how many students received the same emails. Due to limitations on identifiable user information, exact patterns of direct targeting could not be established. It should also be noted that the exact target of each phishing email is unknown. Social engineers could be sending phishing emails to multiple recipients in order to mask a singular intended target. This study is also not aware of how many of the recipients opened the emails.

6.3 Suggestions for Future Research

Future studies on phishing emails can be improved in a variety of ways. To better gauge how the overall university system is targeted, future studies can factor in phishing emails sent to the student users of the university system. Research into student recipient would be beneficial because they might be sent phishing emails for different reasons than university jobholders. Longitudinal data may also provide useful data about certain individuals that are attacked more frequently than others.

The present study's results can be expanded upon in future research to provide even further answers about phishing emails. Future studies can focus more detailed research into the subject of emails and which month they are sent. The "Division" classification of "College" can also be broken down further to offer more detailed information. The "College" can be classified further to show the college departments (Liberal Arts, Engineering, Agriculture, etc.) that have the most recipients targeted for phishing emails. This classification could help explain if certain college departments are targeted more than others for various reasons such as funding or research development. The study could also be expanded is to focus on spear phishing to understand which users are selectively targeted. Finally, the time variables used in this study could also be further used to classify what times of day that phishing campaigns with certain subject content are sent to users. Future studies can compare what day and month certain subject categories are sent to better understand social engineers' attack patterns.

REFERENCES

Blau, P. M. (1964). *Exchange and power in social life*. Transaction Publishers.

Conheady, S. (2014). *Social Engineering in IT Security: Tools, Tactics, and Techniques*. McGraw-Hill Osborne Media.

Cook, K. S., Cheshire, C., Rice, E. R., & Nakagawa, S. (2013). *Social exchange theory* (pp. 61-88). Springer Netherlands.

Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of management*, *31*(6), 874-900.

Deshpande, Nishant. (n.d.) Computer Diseases: Trojan Horses, Viruses & Worms. Retrieved from http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/nd4/article2.html

Dmello, A., Mhatre, G., Lopes, R., & Pen, H. (2013). Spammer Detection by Extracting Message Parameters from Spam Emails. *International Journal of Computer Applications, 78(19).*

Dreeke, Robin. (2011). *It's Not All About Me: The Top Ten Techniques for Building Quick Rapport with Anyone*. People Formula.

Emerson, R. M. (1976). Social exchange theory. *Annual review of sociology*, 335-362.

Hadnagy, Christopher and Fincher, Michele (2015). Phishing Dark Waters: The Offensive and Defensive Sides of Malicious E-mails. Wiley Publishing, Inc.

Hadnagy, Christopher (2011). *Social Engineering: The Art of Human Hacking.* Wiley Publishing, Inc.

Hadnagy, Christopher. (2014). *Unmasking the Social Engineering*. Wiley Publishing Inc.

Hogan, K., & Speakman, J. (2006). *Covert Persuasion*. Audio-Tech Business Book Summaries, Incorporated.

Homans, G. C. (1958). Social behavior as exchange. *American journal of sociology*, 597-606.

Hornecker, E., & Buur, J. (2006, April). Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 437-446). ACM.

Huber, M., Kowalski, S. Nohlberg, M., & Tjoa, S. (2009, August). Towards automating social engineering using social networking sites. In *Computational Science and Engineering, 2009. CSE'09. International Conference on* (Vol. 3, pp. 117-124). IEEE.

Identify and Report Scams and Frauds. (n.d.) https://www.usa.gov/stop-scams-frauds.

Ivaturi, K., & Janczewski, L. (2013). Social Engineering Preparedness of Online Banks: An Asia-Pacific Perspective. *Journal of Global Information Technology Management*, *16*(4), 21-46.

Nelson, A. K. (n.d.). Phishing in 2012 and Beyond.

Perrin, Andrew and Duggan, Maeve. (2015, June). "Americas' Internet Access: 2000-2015. http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/.

Phishing Techniques (n.d.). http://www.phishing.org/phishing-techniques/.

Rusch, J. J. (1999, June). The "social engineering" of Internet fraud. In *Internet Society Annual Conference, http://www. isoc. org/isoc/conferences/inet/99/proceedings/3g/3g_2. htm*.

Skinner, B. F. (1965). *Science and human behavior*. Simon and Schuster.

Splunk (n.d.). What is Machine Data? Retrieved from http://www.splunk.com/content/splunkcom/en_us/resources/machine-data.html.

Statistics Help for Students. (2008). Retrieved from http://statistics-help-for-students.com/

Thompson, S. T. (2013). Helping the hacker? Library information, security, and social engineering. *Information Technology and Libraries*, *25*(4), 222-22.