SOCIAL ENGINEERING ATTACK DETECTION MODEL

A Thesis Presented

By

Yonatan Alemayehu Neda

to

The Faculty of Informatics

of

St. Mary's University

In Partial Fulfillment of the Requirements
for the Degree of Master of Science

July, 2023

ACCEPTANCE

SOCIAL ENGINEERING ATTACK DETECTION MODEL

By

Yonatan Alemayehu Neda

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

____

Internal Examiner

Alembante Mulu Kumlign (PhD)

____

External Examiner

Minale Ashagrie (PhD)

____

Dean, Faculty of Informatics

Alembante Mulu Kumlign (PhD)

July 2023

DECLARATION

This thesis has never been approved for a degree and is not presently being considered by any university.

Unless otherwise mentioned, I declare that the thesis is the result of my own research. I carried out the research on my own, with the supervision and assistance of my research advisor.

Citations with explicit references are used to recognize other sources.

Yonatan Alemayehu

____

Signature

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Asrat Mulatu Beyene (Ph.D.)

____

Signature

Addis Ababa

Ethiopia

July 2023

ACKNOWLEDGEMENTS

I'd like to thank my research adviser, Asrat Mulatu (Ph.D.), for his outgoing advice and support. He has shown me the correct research route and encouraged me to continue with the studies.

This study would not have been possible without the collaboration and valuable opinions of Dashen Bank IT Staffs. Last but not least, I'd want to thank all of my friends for encouraging me to pursue cyber security and for their invaluable assistance in reviewing the customized Model. My greatest gratitude goes to my mother and wife, W/ro. Asrat W/yes and W/ro. Feven Solomon, for their aid and constructive cooperation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF GRAPHS

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

CNN	Convolutional Neural Networks
HR	Human Resource
INSA	Information Network Security Agency
ISO	International Organization for Standardization
IT	Information Technology
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
RNN	Recurrent Neural Networks
SE	Social Engineering
SEA	Social Engineering Attacks
SEADM	Social Engineering Attacks Detection Model
SEADMv2	Social Engineering Attacks Detection Model Version 2
SEDA	Social Engineering Defense Architecture
SIEM	Security Information and Event Management
SMS	Short Message Service
SOC	Security Operation Center
TSEADM	Tailored Social Engineering Attack Detection Model
URL	Uniform Resource Locator

ABSTRACT

Social Engineering is a science of using social interaction as a means to persuade an individual or an organization to comply with a specific request from an attacker where the social interaction, the persuasion or the request involves a computer-related entity. In recent years social engineering attacks have emerged as a growing threat to cyber security, as attackers exploit human vulnerabilities to gain unauthorized access to systems and sensitive information. Ethiopia is no exception, facing an increasing number of such attacks targeted at individuals and institutions.

As a result, multiple Rule-based and machine learning-based models have been developed to address this problem. This thesis proposes a tailored social engineering attack detection model primarily concerned with adapting and modifying SEADM version 2.

The study makes use of survey data, literature review and analysis, and experimentation with real-life scenarios. The results show that the proposed model can be used as a tool by individuals to educate them about the most recent attack technique and to always be vigilant and on the lookout for social engineering attacks. And, has the potential to serve as a valuable tool for organizations and individuals seeking to enhance their cyber security defenses in Ethiopia and similar contexts.

Finally, the study presents a tailored social engineering attack detection model (TSEADM) that has been tested using examples of generalized social engineering attacks, demonstrating that the TSEADM can withstand social engineering attacks.

***Keywords**: Social Engineering, Social Engineering Attacks, Social Engineering Attack Detection Model, Cyber security, Phishing, Rule based Attack Detection Model,*

CHAPTER ONE

INTRODUCTION

Human error is used in social engineering to obtain goods, private information, or access.

These "human hacking" techniques are commonly used in cybercrime to dupe unsuspecting users into disclosing sensitive information, disseminating malware infections, or granting access to restricted systems. These attacks can occur offline, online, or through other interactions.

Social engineering attacks have become a significant threat to organizations and individuals in recent years. Social engineering is a technique used by attackers to manipulate individuals into divulging sensitive information or performing actions that can compromise the security of an organization.

According to a report by Verizon, social engineering attacks accounted for 33% of all data breaches in 2020 (Verizon, 2021). And According to the 2020 Verizon Data Breach Investigations Report, 96% of social engineering attacks enter organizations through email inboxes.

Social engineering poses a threat to the security of all networks, regardless of the strength of their firewalls, cryptographic techniques, intrusion detection systems, and anti-virus software systems. People are more likely to trust other people than computers or other technologies. As a result, they are the weakest link in the security chain.

Human-to-human communication has been improved and accelerated by the advancement of digital communication. Personal and sensitive information, on the other hand, may be available online via social networks and online services that lack security measures to protect this information. Communication systems are vulnerable and can be easily breached by malicious users using social engineering techniques. These attacks are designed to trick individuals or businesses into performing actions that benefit the attackers or providing sensitive data such as social security numbers, health records, and passwords. Social engineering is one of the most difficult challenges in network security because it takes advantage of the natural human tendency to trust.

1.1 STATEMENT OF THE PROBLEM

Social engineering attacks are on the rise in Ethiopia and around the world, due to a lack of information about the subject, which makes people more vulnerable. There is a need for greater research in this sector to aid in raising awareness regarding social engineering attacks.

Awareness and training have been frequently advised as ways for mitigating social engineering attacks because the human being remains a weak point, which social engineers exploit. This thesis tries to address the overarching problem:

A need for a structured and simple method/model that individuals may use to educate themselves and be more vigilant against social engineering attacks by making their emotional state as less exploitable as feasible.

The first aspect to address is researching the current model that depicts all phases and steps of a social engineering attack, according to a standardized definition of social engineering attack.

The second issue is assessing survey data and testing with phishing simulation scenarios to determine elements leading to social engineering attacks and common attack tactics employed by social engineers.

After addressing two of the concerns outlined above, the thesis's underlying problem might be addressed. To accomplish this, the author employs the analyzed data (survey and phishing simulation) and recommendations from social engineering experts.

1.2 RESEARCH QUESTION

This research is motivated to provide an answer to the following questions in the field of social engineering.

1. What is the technique we are using to detect social engineering attacks?
2. What factors contribute for social engineering attacks?
3. What can be done to educate and make people vigilant to social engineering attacks?

1.3 MOTIVATION OF THE RESEARCH

All existing currently used detection strategies are viable options for detecting or preventing social engineering attacks. However, there is still room for improvement, each has weaknesses and must be regularly updated to keep up with the rapid changes.

In addition, because most existing models (Which based on Machine-learning method) rely on the company's expenditures in infrastructure rather than the user/individual (the target of the attack), even huge technological companies that spend a lot of money on security are vulnerable to social engineering attacks.

Therefore, this paper aims at to propose social engineering attack detection models that can be used to teach people how to protect themselves from these types of attacks.

1.4 OBJECTIVES OF THE RESEARCH

GENERAL OBJECTIVE

The aim of this study is to develop social engineering attack detection model

SPECIFIC OBJECTIVES

The main objective of this thesis is:

1. To Identify the factors that expose us to social engineering attack
2. To Develop a tailored social engineering attack detection model which can be utilized as a tool by individuals to educate themselves to be more vigilant and have their guard up always against social engineering attacks
3. To Test and validate the model for detection of social engineering attack

1.5 SCOPE OF THE RESEARCH

This thesis will focus on a social engineering attack detection model based on a rule-based strategy, and the research wanted to focus on developing a model that individuals can use to educate themselves to be more cautious and on guard against social engineering attacks. We also address the following point while doing so:

1. Analyzing different stage of Social Engineering Attack Detection Model and propose a model that it can be used by any individual without requiring specific training on the model itself.

2. Researching the impact of social engineering attacks on individuals and organizations in Ethiopia.
3. Provide standardized social engineering attack example in order to address and test the social engineering attack detection model.

1.6 RESEARCH METHODOLOGY

This study used the survey research design; therefore, we will discuss the data collection and analysis methods used to evaluate the effectiveness of our model. Our goal is to provide a comprehensive and transparent methodology for developing and testing rule-based social engineering attack detection models.

1. DATA GATHERING

   The study use Incident Response report and Phishing Simulation to collect data on social engineering attack, such as the strategies utilized by attackers and the ways they use to deceive victims. These data gathering methods can be used to collect data that can be used to develop rules for detecting social engineering attacks.

   For this study author use a selected bank with the required infrastructure such as firewalls, intrusion detection systems, and security information and event management (SIEM) systems. The selected organization's one-year phishing incident report and the first National Cyber Security Survey report. (INSA, 2014) is analyzed to identify common patterns and trends of the attack.

2. DEFINING RULES

   The study uses the following methods to define the rules for the proposed social engineering attack detection model:

   I. Analyzed Incident response report for social engineering attacks can provide insights into the common tactics employed by social engineers, which can then be used to define the Rule for the detection model.
   II. Experts in social engineering can provide inputs on the various tactics and techniques employed by social engineers, which can be used to define the rules for the detection model.

3. TEST AND VALIDATE

   Once the model is created using a set of defined rules from Analyzed Incident response report and Experts in social engineering, we can test it by using it on a data set of social engineering attacks (real world attack examples).

   Finally, we evaluate the effectiveness of the model based on its ability to detect social engineering attacks accurately and efficiently.

1.7 SIGNIFICANCE OF RESEARCH

This study contributes the following to the field of social engineering:

- Identifying the factors that expose us to social engineering attack through Analyzing Incident response report for social engineering attacks and phishing simulation.
- Provide insights into the common tactics employed by social engineers, which can then be used to define the Rules for the detection model.
- Proposing a tailored social engineering attack detection model which can be utilized as a tool by individuals to educate themselves to be more vigilant and have their guard up always against social engineering attacks.

1.8 LIMITATION OF THE RESEARCH

Most of the organization private or governmental lack a security operation center (SOC) and a SIEM technology, and those that do are afraid to reveal the data and allow a phishing campaign in their firm to test the model.

1.9 ORGANIZATION OF RESEARCH

The following chapters set up this research document: Chapter two is a literature review and related work, which should provide a comprehensive overview of the existing research on the topic, Chapter three is results and findings of data analysis, and the fourth chapter will describe and test the proposed model. The final chapter is the conclusion and recommendation’s part.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORKS

2.1 LITERATURE REVIEW

2.1.1 INTRODUCTION

Social engineering is a type of attack that relies on human psychology to trick victims into giving up sensitive information or taking actions that compromise their security. These attacks can be very effective, as they exploit the natural tendency of people to trust and be helpful. In recent years, there has been a growing interest in developing social engineering attack detection models. These models aim to identify and prevent social engineering attacks by analyzing the behavior of users and the content of messages.

There are a number of different social engineering attack detection models that have been proposed in the literature. Some of these models focus on the linguistic features of messages, while others focus on the behavior of users. Some models are rule-based, while others are machine learning-based.

A recent study by Mouton (2018): proposed a standardized definition for social engineering attack detection models. The study also identified a number of challenges that need to be addressed in order to develop effective social engineering attack detection models.

This literature review will discuss the different types of social engineering attack detection models that have been proposed in the literature. The review will also discuss the challenges that need to be addressed in order to develop effective social engineering attack detection models.

2.1.2 SOCIAL ENGINEERING

Several articles define Social Engineering (SE) and describe a SE attack. The definitions vary and frequently reflect one aspect of an approach relevant to a specific research project. Definitions that are widely accepted and include all of the different entities in SE are required. A large sample of definitions can be found in the literature, and it is abundantly clear that no single formalized definition exists.

The following SE definitions show that there is no single, widely accepted definition:

- "A social engineering attack is one that uses human interaction to gain access to a system or network." Wang, Z., Zhu, H., & Sun, L. (2021). Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods. IEEE
- “Social engineering attacks rely on human psychology to trick victims into giving up confidential information or taking actions that are not in their best interests.” Smith, J. (2023). Social engineering attacks: A guide to prevention and detection. Indianapolis, IN: Wiley
- “Social engineering is the art of manipulating people into giving up confidential information or taking actions that are not in their best interests”. Mitnick, K. D., & Simon, W. L. (2002). The art of deception: Controlling the human element of security. New York, NY: Wiley.
- "Social engineering involves using psychological manipulation to influence individuals or groups to divulge sensitive information or perform an action against their own interests.” IBM Security Intelligence. (https://www.ibm.com/security/learn/social-engineering).
- "Social engineering is the use of deception to manipulate individuals into divulging confidential or personal information that may be used for fraudulent purposes.” Schneier, B. (2002). Beyond fear: Thinking critically about security in an age of terrorism. New York, NY: Copernicus Books.
- "Social engineering is the act of manipulating people into divulging confidential information or performing actions that are not in their best interest.” Sekhar Bhusal, C. (2021). Systematic Review on Social Engineering: Hacking by Manipulating Humans. Journal of Information Security.
- "Social engineering is the art of persuading people to reveal sensitive information.”. Linda Rosencrance. (2021). https://searchsecurity.techtarget.com/definition/social-engineering.

- "Social engineering is a form of cyber-attack that relies on human interaction and often involves tricking people into breaking normal security procedures." Cullen & Armitage. (2016) International Conference on Cyber Security and Protection of Digital Services.
- "Social engineering" is defined as "the use of psychological manipulation to gain access to sensitive information or systems". (2020) https://securityboulevard.com/2020/07/social-engineering-definition-examples-types-of-attacks-and-how-to-prevent-them.
- "Social engineering refers to the use of psychological techniques to trick individuals or organizations into revealing sensitive information or performing actions that are not in their best interest." Titov, D. V, & Filipova, E. E. (2021). Use of neural networks to provide information protection at sensitive facilities of organizations and institutions. Information Security Questions

All of these definitions have one thing in common: a human being is exploited in order to obtain unauthorized information or perform some action. There is no single harmonized definition, as evidenced by the vast array of these definitions.

"Francois Mouton in his research" proposes the following harmonized definitions based on the various definitions given above:

- Social Engineering: The science of using social interaction as a means to persuade an individual or an organization to comply with a specific request from an attacker where the social interaction, the persuasion or the request involves a computer-related entity.
- Social engineer (noun): An individual or group who performs an act of Social Engineering.
- Social engineer (verb2): To perform an act of Social Engineering. When the verb is used in the Past Perfect form, it means a successful Social Engineering attack has occurred. For example, "The target may not know that he or she has been social engineered."
- Social Engineering attack: A Social Engineering attack employs either direct communication or indirect communication, and has a social engineer, a target, a medium, a goal, one or more compliance principles and one or more techniques.

This condensed definition will be used to ensure that the reader understands the terms in the same way that the author does.

2.1.3 SOCIAL ENGINEERING ATTACK

Social engineering attacks (SEA) are currently the most serious cyber security threats. They can be detected but not stopped, according to the researchers. Although social engineering attacks differ, they follow a similar pattern with similar phases.

The common pattern involves four phases:

1) Collect information about the target;

2) Develop relationship with the target;

3) Exploit the available information and execute the attack; and

4) Exit with no traces

Social engineering attacks can be classified into two categories:

- **Human-based**: Human-based attacks are carried out in person by the attacker interacting with the target in order to gather the desired information. As a result, they can only influence a limited number of victims.
- **Computer-based:** To obtain information from targets, software-based attacks use devices such as computers or mobile phones. They can attack a large number of victims in a matter of seconds.

Social engineering attacks can also be classified into three categories, according to how the attack is conducted: social, technical, and physical-based attacks,

- **Social-based attacks** are carried out through relationships with the victims in order to exploit their psychology and emotions. Because they involve human interactions, these are the most dangerous and successful attacks. Baiting and spear phishing are two examples of these attacks.
- **Technical-based attacks** are carried out over the internet via social networks and online service websites in order to obtain desired information such as passwords, credit card information, and security questions.

- **Physical-based attacks** are physical actions taken by the attacker in order to gather information about the target. Such attacks include searching dumpsters for valuable documents.

Human, computer, technical, social and physical-based aspects may be combined in social engineering attacks.

Phishing, impersonation on help desk calls, shoulder surfing, dumpster diving, stealing important documents, diversion theft, fake software, baiting, quid pro quo, pretexting, tailgating, Pop-Up windows, Robocalls, ransom ware, online social engineering, reverse social engineering, and phone social engineering are examples of social engineering attacks.

Through analyzing the different existing classifications of the social engineering attacks, we can also classify these attacks into two main categories: Direct and Indirect.

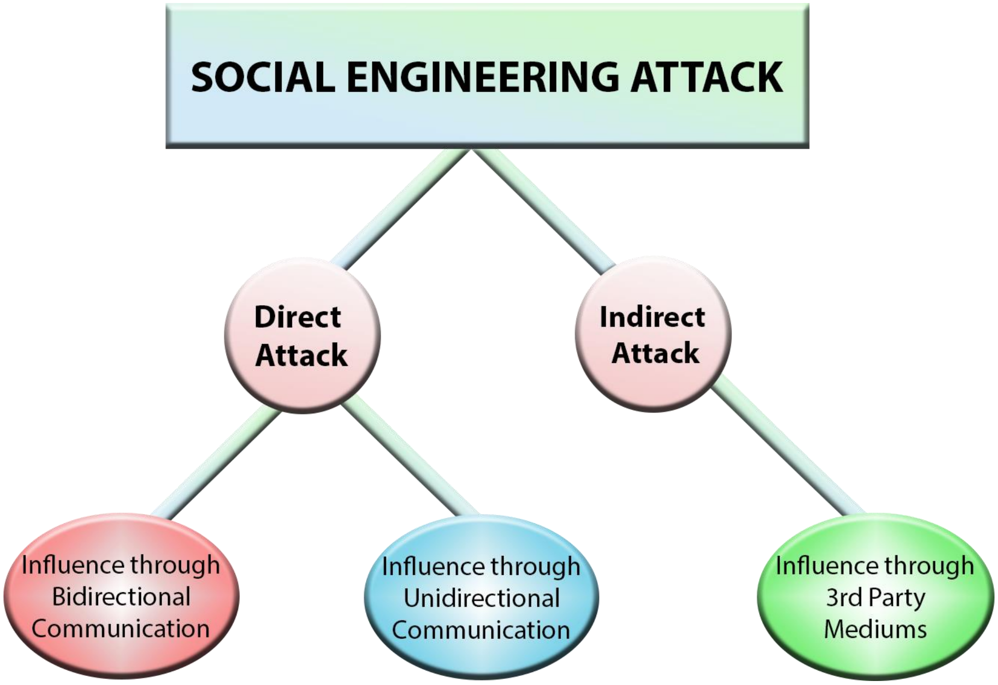

SOCIAL ENGINEERING ATTACK
Direct Attack
Indirect Attack
Influence through Bidirectional Communication
Influence through Unidirectional Communication
Influence through 3rd Party Mediums

Figure 2.1: Social Engineering Attack Classification

Attacks in the first category are carried out through direct contact between the attacker and the victim. They refer to attacks carried out through physical contact, eye contact, or vocal interactions. They may also necessitate the attacker's presence in the victim's working area in order to carry out the attack. Physical access, shoulder surfing, dumpster diving, phone social engineering, pretexting, impersonation on help desk calls, and stealing important documents are examples of these attacks.

As a result, this type of attack is further classified into two modes of communication: bidirectional communication and unidirectional communication.

Bidirectional communication: occurs when two or more parties participate in the conversation, also known as a two-way conversation. Each party is made up of a single person, a group of people, or an organization. An impersonation attack is a common example of this type of attack, in which the social engineer impersonates the target in order to gain access to something that the target has access to.

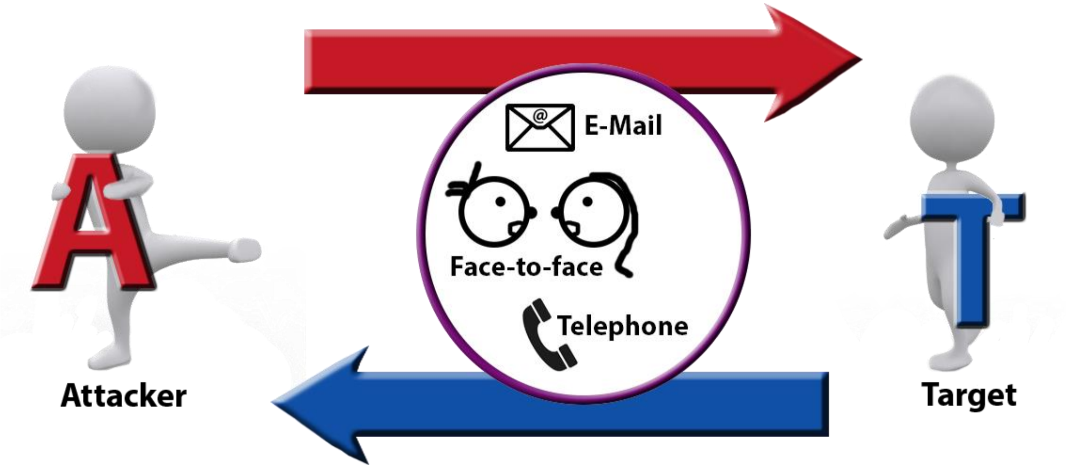


Figure 2.2: Bidirectional Communication

Unidirectional communication is a one-sided conversation in which the social engineer communicates with the target but the target has no way of responding.

Typically, this is accomplished through a communication medium such as bulk e-mails or short message service (SMS). An e-mail phishing attack sent from the attacker to the target is an example of a popular attack in this category.

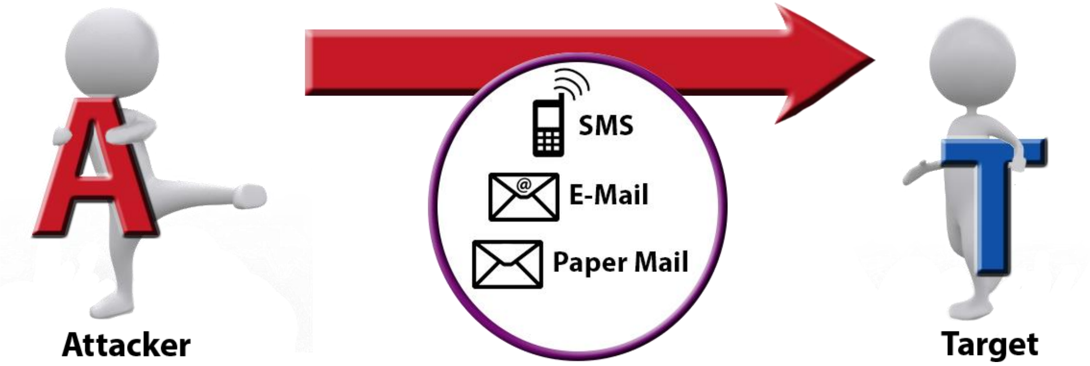


Figure 2.3: Unidirectional Communication

Attacks classified as indirect do not require the presence of the attacker to be launched. The attack can be carried out remotely using malware software delivered via email attachments or SMS messages. These attacks include phishing, fake software, Pop-Up windows, ransom ware, Smishing, online social engineering, and others...

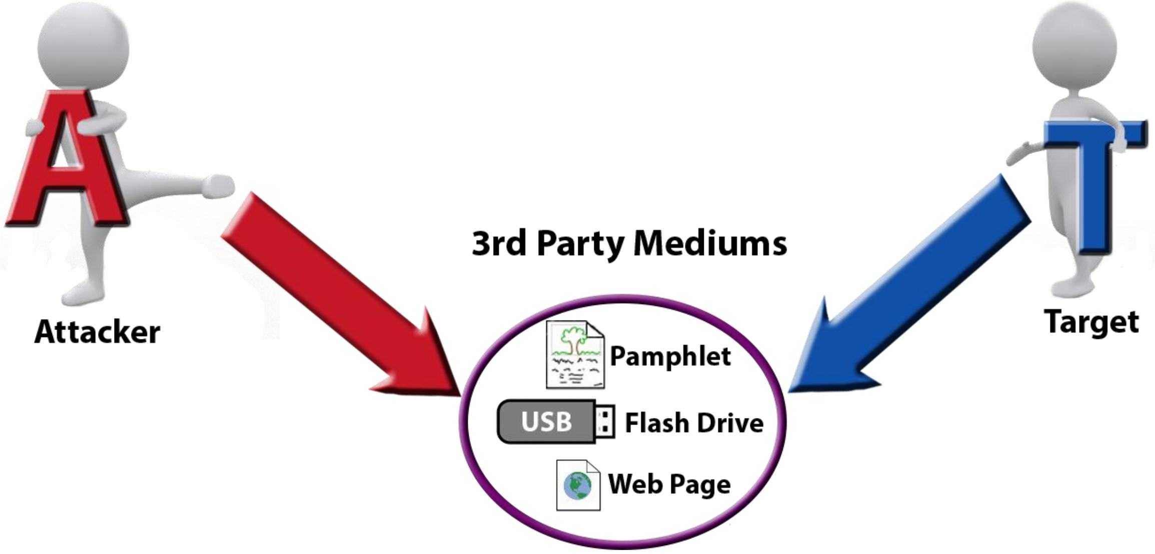


Figure 2.4: Indirect Communication via 3rd Party Medium

2.1.4 SOCIAL ENGINEERING ATTACK DETECTION MODELS

Social engineering attacks are becoming increasingly sophisticated and difficult to detect. Rule-based and machine learning-based models have been developed to address this problem.

Rule-based models rely on a set of predefined rules to identify social engineering attacks. These rules are based on known attack patterns and behaviors. They can be effective in detecting simple attacks, such as phishing and pretexting. However, rule-based models may not be effective in detecting more complex attacks, as the rules may not be comprehensive enough to capture all possible attack scenarios.

Machine learning-based models, on the other hand, use algorithms to learn from data and identify patterns that are indicative of social engineering attacks. These models are more effective in detecting complex attacks, as they can adapt to new attack patterns. Machine learning-based models can also be trained on large datasets, which can improve their accuracy. However, these models require a large amount of data to be effective, and they may be vulnerable to adversarial attacks, where attackers intentionally manipulate the data to evade detection. Several studies have compared the effectiveness of rule-based and machine learning-based models in detecting social engineering attacks. One study found that machine learning-based models outperformed rule-based models in detecting phishing attacks. Another study found that both rule-based and machine learning-based models were effective in detecting pretexting attacks, but machine learning-based models had higher accuracy and lower false positive rates.

The SEADMs listed below assist individuals in self-questioning, evaluating, and protecting themselves from various types of social engineering attacks. Social engineering is extremely difficult to detect because the social engineer possesses a wide range of skills and effective techniques that prey on human vulnerabilities, allowing these attacks to go undetected for long periods of time. Making detection even more difficult is the fact that many people are unaware of this technique's potential threat and dire Disadvantage sequences for individuals and institutions.

2.1.4.1 THE SOCIAL ENGINEERING ATTACK DETECTION MODEL (SEADM)

It provides a clear guideline for determining whether an individual is the victim of a social engineering attack. It accomplishes this by presenting a set of binary states in a diagram, as shown in Figure 1. The user must proceed through the diagram until they reach the end state. The ending state will assist the user in determining whether to grant the requester access or elevate the requester's request.

SEADM provides a detailed description of each model state as well as real world scenarios describing how the model could be used to detect social engineering attacks.

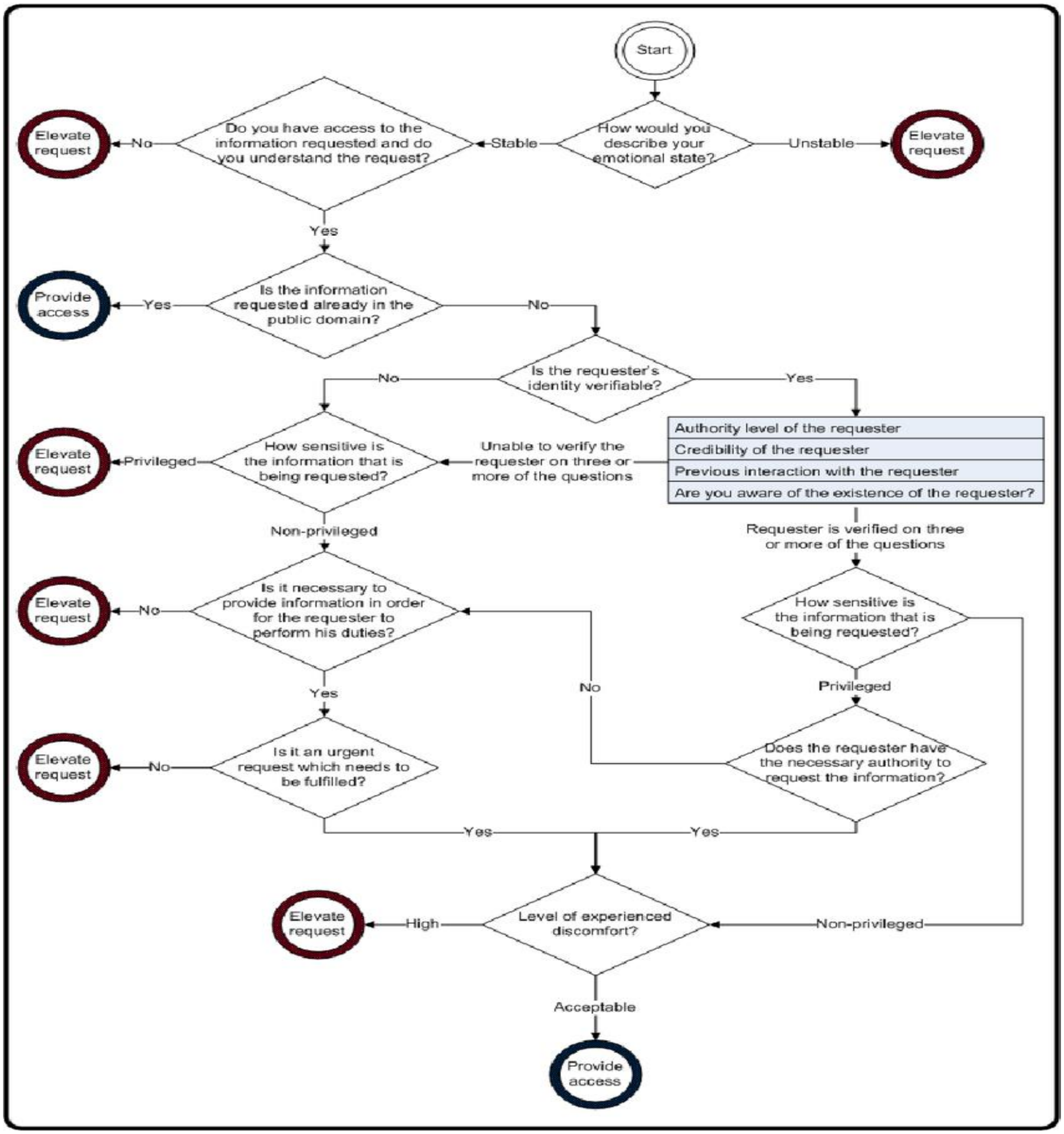

Start
How would you describe your emotional state?
Stable
Unstable
Elevate request
Do you have access to the information requested and do you understand the request?
No
Elevate request
Yes
Is the information requested already in the public domain?
Yes
Provide access
No
Is the requester's identity verifiable?
No
Yes
Authority level of the requester
Credibility of the requester
Previous interaction with the requester
Are you aware of the existence of the requester?
Unable to verify the requester on three or more of the questions
Requester is verified on three or more of the questions
How sensitive is the information that is being requested?
Privileged
Elevate request
Non-privileged
Is it necessary to provide information in order for the requester to perform his duties?
No
Elevate request
Yes
Is it an urgent request which needs to be fulfilled?
No
Elevate request
How sensitive is the information that is being requested?
Privileged
Does the requester have the necessary authority to request the information?
No
Yes
Yes
Non-privileged
Level of experienced discomfort?
High
Elevate request
Acceptable
Provide access

Figure 2.5: Social Engineering Attack Detection Model (Bezuidenhout, Mouton & Venter, August 2010)

2.1.4.2 SOCIAL ENGINEERING ATTACK DETECTION MODEL VERSION 2 (SEADMV2)

The Social Engineering Attack Detection Model has been updated (SEADMv2). Figure 2 depicts a diagram of this model.

Figure 2 contains more states than Figure 1, and the states are color coded as well. The colors used for the states are used to distinguish between the various types of states supported by the model. Request (yellow) states are those that deal with the request itself. Blue states are receiver states, and they are concerned with whether or not a person understands what is being requested. The green states are concerned with the requester and any information that can be determined about the requester, whereas the red states are concerned with whether the requester can be verified using a third party.

While the SEADMv2 states refer to four different parties, it is important to remember that the SEADMv2 will still be used by an individual receiving a request. This model will be used by the individual to determine whether the request can be completed or if it should be deferred or referred to someone else.

Each state in the SEADMv2 is thoroughly described, and three scenarios are provided to clearly demonstrate how it can be used in practice. The SEADMv2 can detect bidirectional, unidirectional, and indirect communication. Finally, it is worth noting that the states dealing with the receiver determining his or her emotional state and level of discomfort are missing from the SEADMv2.

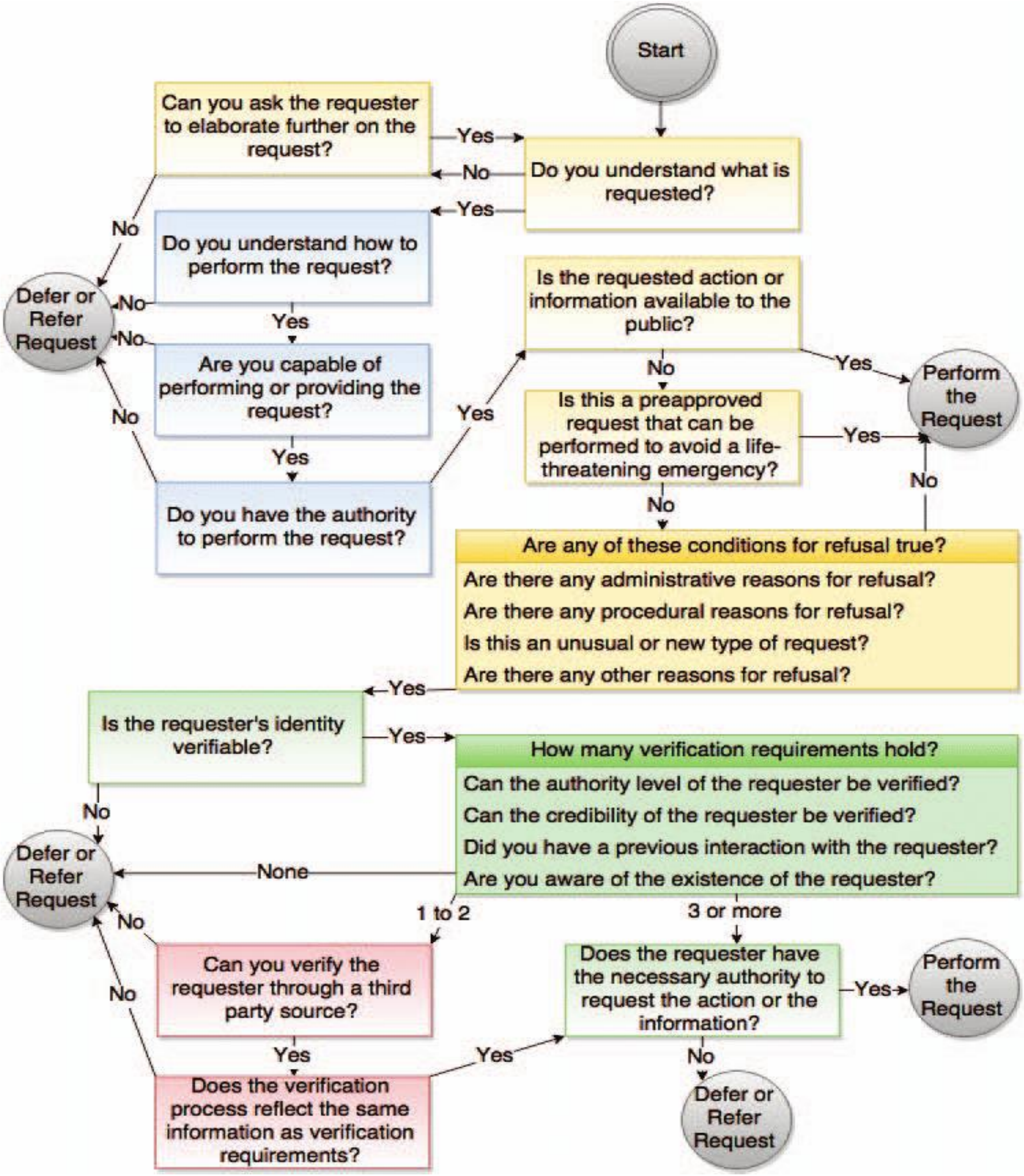

Start
Do you understand what is requested?
Can you ask the requester to elaborate further on the request?
Yes
No
Yes
No
Do you understand how to perform the request?
Defer or Refer Request
No
Yes
No
Are you capable of performing or providing the request?
No
Yes
Do you have the authority to perform the request?
Yes
Is the requested action or information available to the public?
Yes
Perform the Request
No
Is this a preapproved request that can be performed to avoid a life-threatening emergency?
Yes
No
No
Are any of these conditions for refusal true?
Are there any administrative reasons for refusal?
Are there any procedural reasons for refusal?
Is this an unusual or new type of request?
Are there any other reasons for refusal?
Yes
Is the requester's identity verifiable?
Yes
How many verification requirements hold?
Can the authority level of the requester be verified?
Can the credibility of the requester be verified?
Did you have a previous interaction with the requester?
Are you aware of the existence of the requester?
No
Defer or Refer Request
None
1 to 2
3 or more
No
No
Can you verify the requester through a third party source?
Yes
Does the verification process reflect the same information as verification requirements?
Yes
Does the requester have the necessary authority to request the action or the information?
Yes
Perform the Request
No
Defer or Refer Request

Figure 2.6: Social Engineering Attack Detection Model version 2(Francois Mouton, November 2018)

2.1.3.3 OTHETR SOCIAL ENGINEERING DETECTION MODELS

The Social Engineering Attack Detection Model (SEADM) is not the only method for identifying social engineering attacks. Other detection models used to detect or prevent social engineering attacks will be discussed in this section.

I. **Social Engineering Detection Using Neural Networks**: A neural network is a computational model that is biologically inspired by brain neurons. Neural networks employ various layers of nodes that are trained with training data to produce the correct output given valid input values.

II. **Social Engineering Defense Architecture (SEDA):** is a software system that detects telephonic social engineering attacks. The SEDA proposes using a voice signature authentication system. The idea is to link the voice signatures to a database of personal information about the employees, such as their name, corporate affiliation, job title, and all phone numbers from which a specific employee might call.

III. **Detection of Social Engineering Attacks through Natural Language Processing**: It is similar to the SEDA in that it uses a software system to detect social engineering attacks and uses natural language processing to detect them. This mechanism, however, only detects textual social engineering attacks such as phishing emails. Through a series of steps, the software system detects social engineering attacks using natural language processing.

2.2 RELATED WORKS

Social engineering attacks are a growing threat to information security, and detecting such attacks is a challenging task. Researchers have proposed various models to detect social engineering attacks, and several articles have been published on this topic.

One such article by M. Saadatmand and M. Zamani (2018)."A Review of Social Engineering Attack Detection Models". Provides an overview of various social engineering attack detection models, including rule-based, machine learning- based, and hybrid models. The article also discusses the strengths and limitations of each model and identifies areas for future research.

Ahmed AlEroud. (2021)."Detecting Social Engineering Attacks with Natural Language Processing and Machine Learning”. Proposes a model for detecting social engineering attacks in online communication using natural language processing and machine learning techniques. The authors demonstrate the effectiveness of their model using a dataset of email messages.

Another article on this topic by A. Singh and S. Bhattacharya (2020), "Detecting Social Engineering Attacks using Machine Learning Techniques”. Proposes a machine learning-based approach to detect social engineering attacks by analyzing the behavior of users on social media platforms. The authors used various machine learning algorithms to analyze user behavior patterns and identify potential social engineering attacks.

Zainab Alkazemi (2021). "A Hybrid Model for Social Engineering Detection using Machine Learning and Rule-Based Approaches”. This article presents a hybrid model that combines machine learning and rule-based approaches to detect social engineering attacks. The authors evaluate their model using a dataset of phishing emails and show that it outperforms several baseline models.

Yanbin Lu and colleagues. (2020). "Detecting Social Engineering Attacks on Social Media using Deep Learning,”. This article proposes a deep learning model for detecting social engineering attacks on social media platforms. The authors use a dataset of tweets and demonstrate the effectiveness of their model in detecting phishing attacks and scams.

S. S. Ahmed and A. S. Alshomrani (2019). "A Framework for Detecting Social Engineering Attacks using Machine Learning Techniques”. Proposed a framework for detecting social engineering attacks using machine learning techniques. The framework uses various features such as the user's browsing behavior, typing behavior, and mouse movements to detect social engineering attacks. The authors used various machine learning algorithms to analyze the features and identify potential social engineering attacks.

These articles provide valuable insights into the different approaches and techniques used to detect social engineering attacks. They highlight the importance of developing effective detection models to improve information security and protect against social engineering attacks.

Overall, these state-of-the-art social engineering attack detection models demonstrate the importance of using advanced techniques and approaches to detect attacks and improve information security. These models are continuously evolving, and researchers are expected to develop more sophisticated models in the coming years to keep up with the ever-changing threat landscape.

In recent years, social engineering attacks have become increasingly prevalent and sophisticated**,** posing a significant threat to individuals, organizations, and society as a whole. As a result, researchers have developed several state-of-the-art models for detecting social engineering attacks.

One of the most promising approaches is the use of machine learning algorithms to identify patterns and anomalies in communication patterns. Researchers have used various features such as the language used in messages, the timing of messages, and the relationship between senders and receivers to develop models that can detect social engineering attacks with high accuracy.

Another approach is to use natural language processing (NLP) techniques to analyze the content of messages and identify specific linguistic indicators of social engineering attacks. For instance, researchers have used sentiment analysis, topic modeling, and named-entity recognition to identify suspicious messages that may be part of a social engineering attack.

Moreover, some researchers have proposed the use of graph-based models to detect social engineering attacks. These models use graph theory to represent the communication patterns between individuals and detect anomalies in the network that may indicate a social engineering attack.

Here are some state-of-the-art models that have been developed:

1. **Deep Learning-Based Models:**
   Deep learning-based models have gained popularity for social engineering attack detection. These models use advanced neural networks to learn from the features extracted from social engineering attacks and detect them. In particular, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been used for this purpose.
2. **Hybrid Models:**
   Hybrid models that combine rule-based and machine learning- based approaches have been developed to improve detection accuracy. These models use predefined rules to identify suspicious behavior and then apply machine learning techniques to classify such behavior as social engineering attacks or normal user behavior.

3. **User Behavioral Models:**

   User behavioral models use the behavior of users to detect social engineering attacks. These models analyze user behavior patterns such as mouse movements, typing speed, and browsing behavior to detect anomalies that may indicate an attack.

4. **Natural Language Processing (NLP)-**

   Based Models: NLP-based models have been developed to detect social engineering attacks in emails, chat messages, and other forms of text-based communication. These models analyze the language used in the messages and identify suspicious patterns that may indicate an attack.

5. **Data-Driven Models:**

   Data-driven models use large datasets to learn from social engineering attacks and improve detection accuracy. These models use techniques such as clustering and association rule mining to identify patterns and detect attacks.

Overall, the state-of-the-art models developed in the last five years have shown promising results in detecting social engineering attacks.

However, there is still room for improvement, particularly in addressing the issue of false positives and false negatives. Additionally, as social engineering attacks continue to evolve, it is crucial for researchers to keep developing and testing new models to stay ahead of the attackers.

2.2.1 MODELS COMPARATIVE ANALYSIS

Social engineering attack detection models aim to identify and prevent attacks that manipulate individuals into divulging confidential information or performing actions that compromise security. These models can be classified into two categories: rule-based and machine learning-based models.

Based on accuracy, scalability, training and education capabilities, attack scope coverage, resource requirements, and usability, the following are some pros and cons of each approach based on their categories:

I. Rule-based models:

Strengths:

- They are simple to design and implement.
- They can be effective at detecting known social engineering attack patterns.
- They can be easily updated to include new attack patterns.

Weakness:

- They may not be effective at detecting new or unknown attack patterns.
- They may generate false positives or false negatives.
- They may require frequent updates and maintenance.

II. Machine learning-based models:

Strengths:

- They can detect new or unknown attack patterns.
- They can learn from historical data and improve over time.
- They can reduce false positives and false negatives with proper training.

Weakness:

- They require large amounts of high-quality training data.
- They can be computationally intensive and require significant resources.
- They can be susceptible to adversarial attacks.

The model categories listed above can be utilized for speech, text, or both attack detection. Voice and text-based social engineering attack detection models each have strengths and weakness. Each strategy is summarized below:

A. Voice-based social engineering attack detection models:

Strengths:

- Voice biometrics can provide accurate identification of individuals and detect Impersonation attempts.
- Natural language processing (NLP) can analyze the content and context of a conversation to identify suspicious patterns or language cues that indicate a social engineering attack.
- Behavioral analysis can detect anomalies in a user's behavior that may indicate a social engineering attack.

Weakness:

- Voice biometrics may not be effective if the attacker can imitate the victim's voice.
- NLP may generate false positives or false negatives if the context is not fully understood.
- Behavioral analysis may not detect attacks that are well-planned and executed.

B. Text-based social engineering attack detection models:

Strengths:

- Machine learning algorithms can analyze large amounts of data to identify patterns and anomalies that may indicate a social engineering attack.
- NLP can analyze the content and context of an email or text message to identify suspicious patterns or language cues that indicate a social engineering attack.
- URL analysis can detect malicious links in emails or text messages and prevent users from clicking on them.

Weakness:

- Machine learning algorithms may not be effective if the training data is not representative of the actual attack patterns.
- NLP may generate false positives or false negatives if the context is not fully understood.
- URL analysis may not detect attacks that use legitimate links or redirect users to malicious sites.

The table below summarizes the comparison of social engineering attack detection models:

Table 2.1: Comparison of Social Engineering Detection Models

	SEADM	SEADM Psychological Measure	SEADMv2	Neural Networks	SEDA	Natural Language Processing
Detection Model Used	States	Cognitive functioning measures	States	Neural net works	Voice signatures	Natural language processing
User Interaction Required	Yes	Yes	Yes	Yes	No	No
Types of Social Engineering Attacks Which Model Can Detect	Textual and verbal	Textual and verbal	Textual and verbal	Textual and verbal	Verbal	Textual
Strengths:	Modular design.	Quick to perform tests. Provide a concrete way of determining emotional state.	Color codes to differentiate types of states. More state transitions than the SEADM. More modular design than the SEADM. Caters for bidirectional, unidirectional and indirect communication.	Accurate at detecting attacks.	No user inter action required. Prevents same social engineer targeting different employees.	No user interaction required. Processes text rapidly. Accurate at detecting attacks.
Weakness:	Requires user to determine own emotional state. Only caters for bidirectional communication.	Tests could become repetitive if performed too many times.	No states to examine the emotional state of the user.	Never been tested in a real-world scenario. Tedious for the user to enter values into the input nodes.	Social engineer Could trick the system by using voice recordings of the person they are imitating. Only works for verbal social engineering attacks.	Only works for Textual social engineering attacks.

In conclusion, as demonstrated by the comparison table summary above, all detection techniques are valid solutions for detecting or avoiding social engineering attacks. Rule-based models are simple and effective at detecting known social engineering attacks, but they may not be sufficient to handle new or unknown attacks. Machine learning-based models can adapt to new attack patterns and improve over time, but they require significant resources and training data. And Voice-based models are effective at detecting voice-based social engineering attacks, but may not be able to detect attacks that are well-planned and executed. Text-based models are effective at detecting phishing emails or text messages, but may not be able to detect attacks that use legitimate links or redirect users to malicious sites. The best solution for any organization will depend on its specific needs, resources, and risk profile/tolerance.

Therefore, this paper aims at to propose social engineering attack detection models that can be used to teach people how to protect themselves from these types of attacks without any technical knowledge required. Unlike most of existing models which rely on the company's expenditures in infrastructure rather than the user/individual (the target of the attack), that’s why even huge technological companies that spend a lot of money on security are vulnerable to social engineering attacks.

And the author believes that the most effective way to detect social engineering attacks is to combine several of these models (those that are simple to implement and update and aim to teach individuals and make them more vigilant to social engineering attacks) such as the "Think Before You Click" model, the "Trust but Verify" model, and the "Education and Awareness" model and tailor them to the specific needs and vulnerabilities of the people being targeted for.

CHAPTER THREE

DATA ANALYSIS, FINDINGS AND DISCUSSION

3.1 INTRODUCTION

This chapter states the model's data analysis, design, and development. The findings are provided in accordance with the defined methodology, and study objectives, in order to answer the research questions. The findings in this section are based on data analysis.

3.2 INCIDENT RESPONSE REPORT ANALYSIS

The study makes use of Dashen Bank one-year (2022/23) phishing incidence record and the first INSA's 2014 national cyber security survey report.

3.2.1 NATIONAL CYBER SECURITY SURVEY REPORT (INSA's 2014)

To demonstrate Ethiopia's current vulnerability/ and factors that contributes to cyber security in general and social engineering attacks in particular, the study analyzes a survey report published by Ethiopia's information network security agency in 2014 EC. The survey's goal was to assess the current state of cyber security at the nation's key infrastructures and institutes.

And the survey is conducted using the following tools:

- ✓ Critical mass cyber security requirement standards
- ✓ OWASP
- ✓ NIST framework
- ✓ ISO 27001

This survey is carried out in the country's key infrastructures and institutes, which are classified into six sectors based on their overall impact.

1. Media Institutes
2. Financial Institutes
3. Institute for Critical Infrastructure

4. Educational establishment
5. Institute for Peace and Security
6. Institute of Trade and Markets

The assessment is performed on the selected institute using the aforementioned tools and five key measuring criteria:

1. Policies and Procedures
2. Infrastructure Development
3. Capacity Building
4. Awareness and Culture
5. Leadership and Management Standard

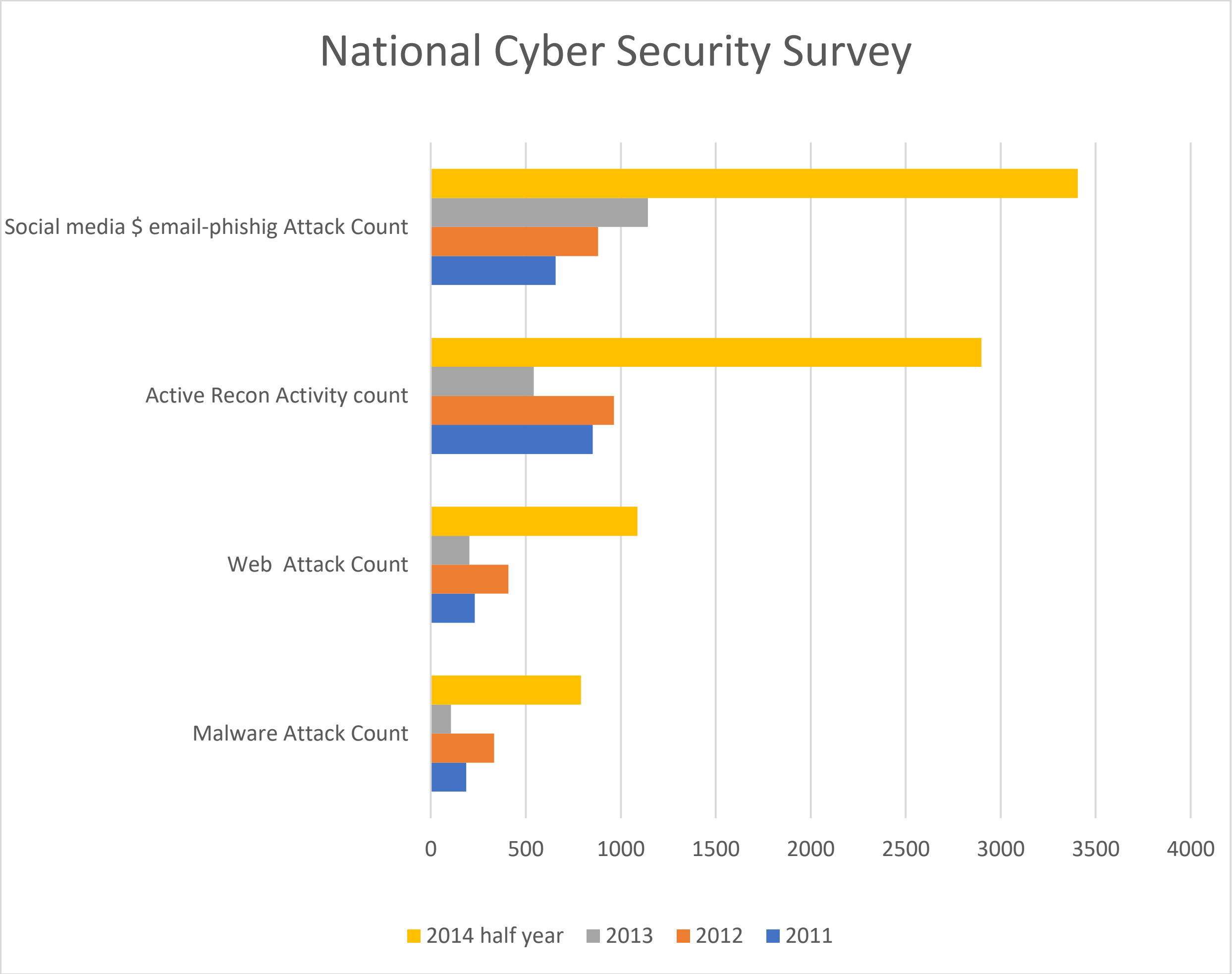
It has been observed that the surveyed organization's security policies and procedures, technology, training, Awareness and Culture regarding cyber security in the country are low, and employees ranging from top level management to ordinary workforce of the key institutes surveyed on mentioned criteria are ranked low. And, as a result of this lack of security policies and procedures, technology, training, awareness and culture on these key countries' infrastructure and institutions are on the rise.

The survey result is summarized in the below:

- ✓ Institutes such as trade and transportation, health, education, and the minister's office have **low** scores, ranging from 0 to 50%.
- ✓ Financial institutes and peace and security institutes have a **medium** score, which means they are in the 50 - 75% range.
- ✓ No Ethiopian key institute scored **high** or above 75% in this survey.

According to the survey report, the country's cyber security is at high risk. As a result, if this high risk is not minimized or handled as quickly as feasible, Ethiopia would suffer economic, political, and social difficulties.

The overall national cyber security survey for the year 2011, 2012, 2013 and 2014 is summarized in the Graph below.



Graph 3.1 Comparison of various cyber-attacks over the last four years (INSA Survey, 2014)

As we can see from the summery chart, the attack rate is rapidly increasing at the moment, with the majority of attacks employing social engineering techniques to gain access and progress to the next phase.

As a result, having a social engineering attack detection model or educating society about social engineering attacks is necessary to reduce the danger, which is growing rapidly.

3.2.2 DASHEN BANK INCIDENT RESPONSE REPORT

Dashen Bank is a leading commercial bank in Ethiopia. In recent years, banks are experiencing a significant increase in phishing attacks. These attacks targeted both employees and customers. The bank's incident response team was able to successfully mitigate the majority of these attacks.

This paper examines the bank's response to phishing incidents in 2022-23. Emails that seemed to be from reputable sources, such as from the bank itself, known service providers or a government body, were often used in the attacks. The emails would frequently contain links or attachments that, when clicked, would infect the victim's machine with malware or redirect the user to an unsafe page.

During the stated year, over 5752 security incidents were reported, with 3095 of these security incidents being phishing incident reports, accounting for 53.8% of all security incidents documented in the bank.

From the 3095 phishing reports, 398 events are seen for first time, implying that while one phishing email would be delivered to numerous employees, the true number is 398.

From the initially spotted phishing attack, 171 are rated high in severity, accounting for 43% of the attacks performed.

When we examine the attacks, particularly those rated High in severity, we see that the attack primarily adopts the following technique to exploit and gain what they desire:

1. Emails and websites links that look like they're from a legitimate source.
2. The Emails contain a sense of urgency, such as a warning if you don’t act immediately.
3. The Emails have attachments with payloads (more than 20%).
4. The Emails have malicious link (up to 80%).

3.3 PHISHING SIMULATION

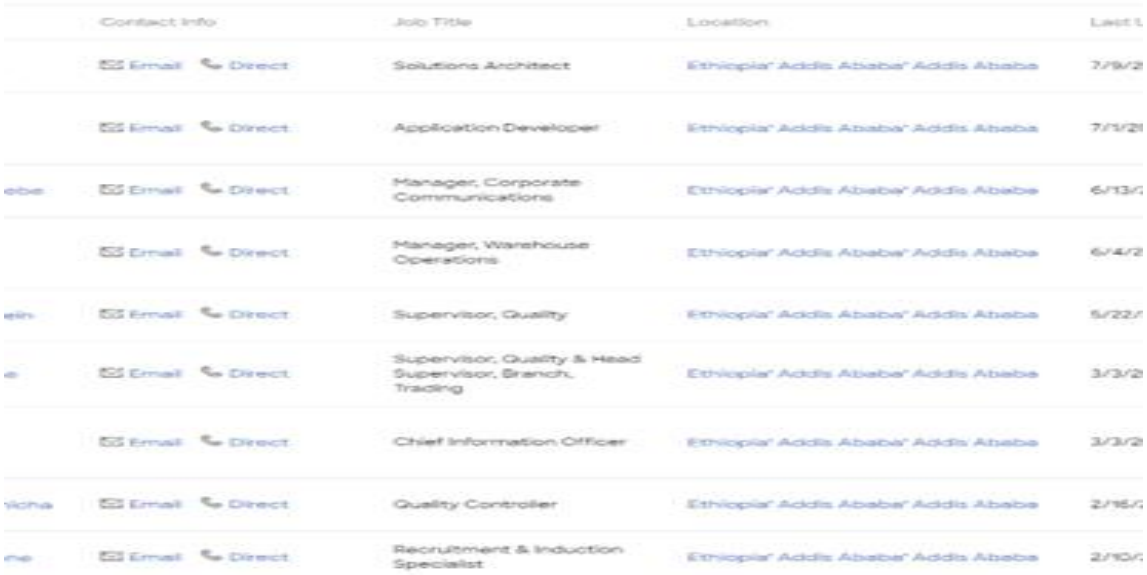
Phishing simulation is a security awareness training technique that uses simulated phishing emails to test employees' ability to identify and report phishing attacks. Phishing simulation is an effective way to improve employee security awareness and reduce the risk of phishing attacks.

When employees are regularly exposed to simulated phishing emails, they become more familiar with the signs of a phishing attack and are more likely to report suspicious emails to IT security.

As a result, in order to acquire information about individuals' abilities to identify and report phishing attacks, the study employs a successful social engineering examination on employees. According to the findings, employees are extremely vulnerable to social engineering attacks, which will be used by the attacker to compromise the user's account.

The Social engineering exploitation used, follows different phases:

1. User E-mail Enumeration



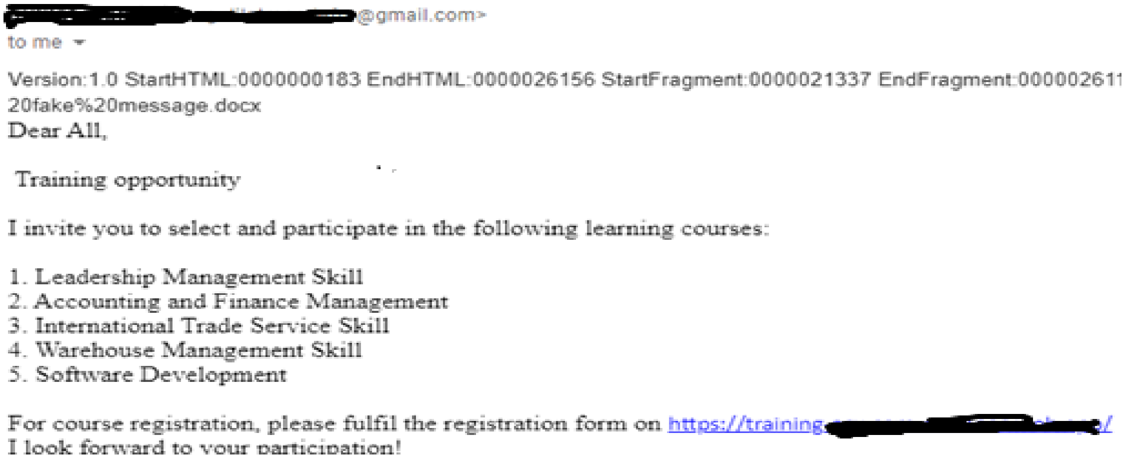
	Contact Info	Job Title	Location	Last U
	✉ Email ☎ Direct	Solutions Architect	Ethiopia' Addis Ababa' Addis Ababa	7/9/2
	✉ Email ☎ Direct	Application Developer	Ethiopia' Addis Ababa' Addis Ababa	7/1/2
ebe	✉ Email ☎ Direct	Manager, Corporate Communications	Ethiopia' Addis Ababa' Addis Ababa	6/13/
	✉ Email ☎ Direct	Manager, Warehouse Operations	Ethiopia' Addis Ababa' Addis Ababa	6/4/2
ein	✉ Email ☎ Direct	Supervisor, Quality	Ethiopia' Addis Ababa' Addis Ababa	5/22/
e	✉ Email ☎ Direct	Supervisor, Quality & Head Supervisor, Branch, Trading	Ethiopia' Addis Ababa' Addis Ababa	3/3/2
	✉ Email ☎ Direct	Chief Information Officer	Ethiopia' Addis Ababa' Addis Ababa	3/3/2
icha	✉ Email ☎ Direct	Quality Controller	Ethiopia' Addis Ababa' Addis Ababa	2/16/
ne	✉ Email ☎ Direct	Recruitment & Induction Specialist	Ethiopia' Addis Ababa' Addis Ababa	2/10/

Fig 3.1 User Email Enumeration Sample Data

2. Developing pretexting

   Is a type of social engineering in which an auditor attempts to persuade staff to provide user account information while gaining access to systems? The scam artists create a tale or pretext to deceive the victim in this type of attack.

This is our pretexting for the selected users:



[redacted]@gmail.com>
to me ▾

Version:1.0 StartHTML:0000000183 EndHTML:0000026156 StartFragment:0000021337 EndFragment:00000261120fake%20message.docx
Dear All,

Training opportunity

I invite you to select and participate in the following learning courses:

1. Leadership Management Skill
2. Accounting and Finance Management
3. International Trade Service Skill
4. Warehouse Management Skill
5. Software Development

For course registration, please fulfil the registration form on https://training[redacted]/
I look forward to your participation!

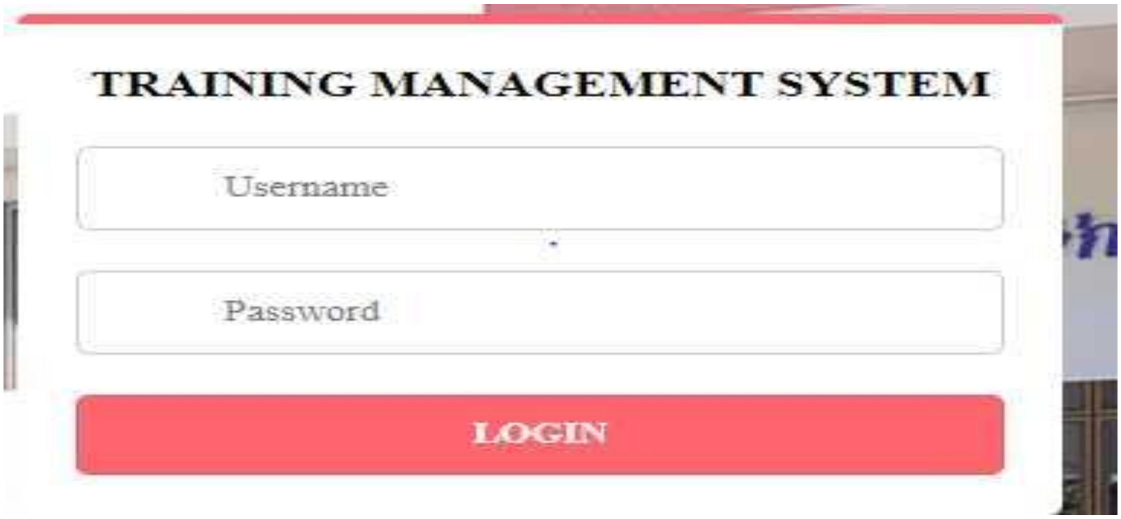
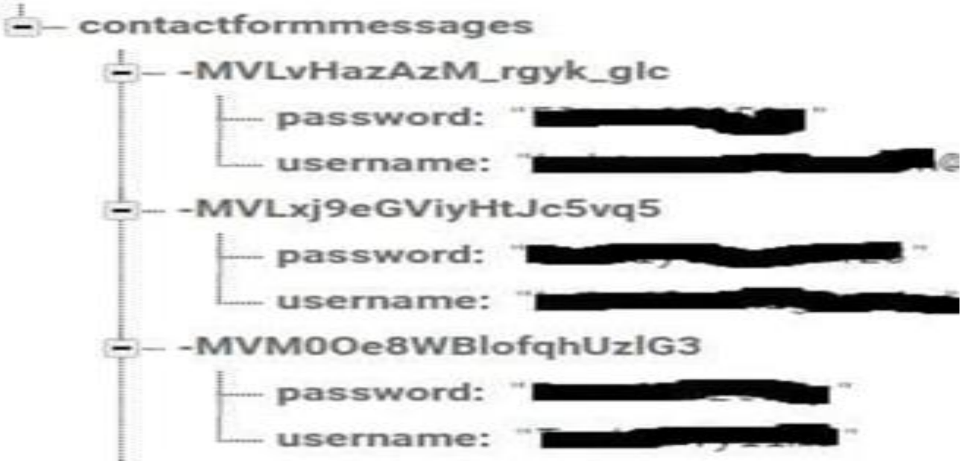


Fig 3.2 Pretexting Email Message and Attached Link Landing Page

3. Collect successful results

   Here we provide some of the user account that are exploited using social engineering technique. Out of 30 Employees selected, we got 7 successful results.



```
contactformmessages
    -MVLvHazAzM_rgyk_glc
        password: "[redacted]"
        username: "[redacted]
    -MVLxj9eGViyHtJc5vq5
        password: "[redacted]"
        username: "[redacted]
    -MVM0Oe8WBlofqhUzlG3
        password: "[redacted]"
        username: "[redacted]"
```

Fig 3.3 Sample Successful Result Collected

- ❖ Therefore, we can conclude that Employees are vulnerable to social engineering attack, and the root causes for this is that Employers didn’t get a proper cyber security awareness training depending on their responsibilities.

3.4 FACTORS CONTRIBUTING TO SOCIAL ENGINEERING ATTACKS

Based on our analyzed data, we can conclude that social engineering attacks are successful because they take advantage of human psychology. Attackers use a variety of tactics to deceive victims into exposing sensitive information or acting in ways that put them or their organizations in danger.

The study found the following as the most common factors that contribute to social engineering attacks:

3.4.1 Lack of awareness

Many people are not aware of the different social engineering techniques that attackers use. This lack of awareness makes them more likely to fall victim to an attack.

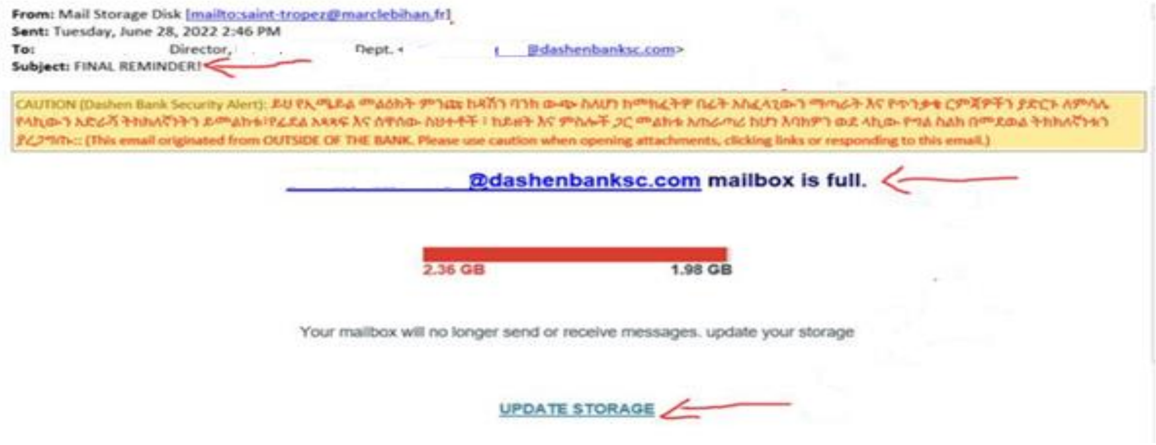
According to the survey report's awareness and culture measurement criteria, this is represented in the national cyber security survey report (INSA, 2014);

- ✓ Institutes such as trade and transportation, health, education, and the minister's office have low scores, ranging from 0 to 50%.
- ✓ Financial institutes and peace and security institutes have a medium score, which means they are in the 50 - 75% range.
- ✓ No Ethiopian key institute scored high or above 75% in this survey.

3.4.2 Urgency

Attackers often create a sense of urgency in their requests, which can make victims feel pressured to comply. This pressure can lead to victims making poor decisions, such as clicking on a malicious link or providing sensitive information.

Example from analyzed attacks which use sense of Urgency to trick individual to click on malicious link.



From: Mail Storage Disk [mailto:saint-tropez@marclebihan.fr]
Sent: Tuesday, June 28, 2022 2:46 PM
To: Director, Dept. < @dashenbanksc.com>
Subject: FINAL REMINDER!

CAUTION (Dashen Bank Security Alert): ይህ የኢሜይል መልዕክት ምንጩ ከዳሸን ባንክ ውጭ ስለሆነ ከመክፈትዎ በፊት አስፈላጊውን ማጣሪያ እና የጥንቃቄ እርምጃዎችን ያድርጉ ለምሳሌ የላኪውን አድራሻ ትክክለኛነትን ይመልከቱ፣የፊደል አጻጻፍ እና ሰዋሰው ስህተቶች ፣ ከይዘት እና ምስሎች ጋር መልክቱ አከራካሪ ከሆነ እባክዎን ወደ ላኪው የግል ስልክ በመደወል ትክክለኛነቱን ያረጋግጡ:: (This email originated from OUTSIDE OF THE BANK. Please use caution when opening attachments, clicking links or responding to this email.)

@dashenbanksc.com mailbox is full.

Your mailbox will no longer send or receive messages. update your storage

Fig 3.4 Phishing Attack Example Which Use Sense of Urgency

3.4.3 Trust

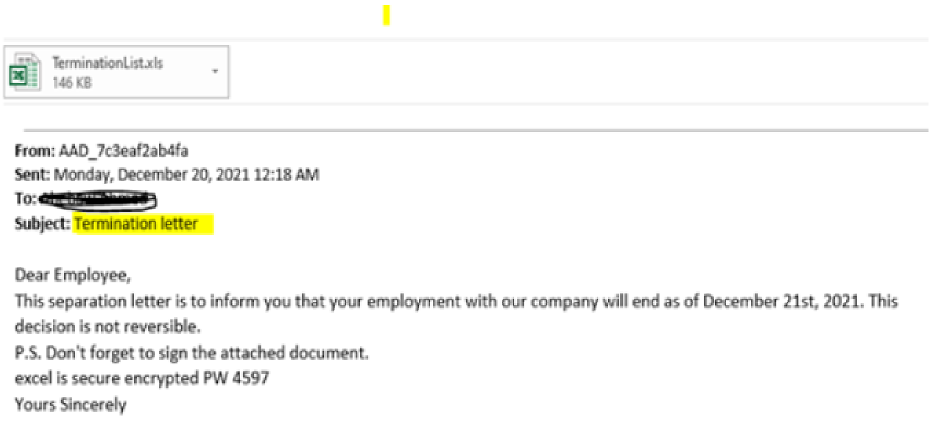
Attackers often try to build trust with their victims. They may pose as a legitimate employee, customer, or vendor. This trust can make victims more likely to believe the attacker and take action.

For example, our pretexting in the simulated phishing was developed by copying a user who is generally in charge of inviting employees for training, and those six folks became victims just by observing the individual's name and believing it was authentic.

3.4.4 Fear and Curiosity

Both are emotion that can be exploited by social engineers. Attackers may use curiosity to trick their victims into clicking on malicious links or opening attachments that contain malware or use fear to manipulate their victims.

Example from analyzed attacks which use emotions like fear and curiosity to trick individual to click on malicious link.



TerminationList.xls
146 KB

From: AAD_7c3eaf2ab4fa
Sent: Monday, December 20, 2021 12:18 AM
To: [redacted]
Subject: Termination letter

Dear Employee,
This separation letter is to inform you that your employment with our company will end as of December 21st, 2021. This decision is not reversible.
P.S. Don't forget to sign the attached document.
excel is secure encrypted PW 4597
Yours Sincerely

Fig 3.5 Phishing Attack Example Which Use Fear and Curiosity Emotion to Trick Individual.

3.4.5 Lack of security controls

Organizations that do not have strong security controls in place are more vulnerable to social engineering attacks. These controls can help to protect organizations from attackers by making it more difficult for them to gain access to sensitive information or systems. The study recommends the following key points:

- ✓ Develops and implements cyber security strategies and policies;
- ✓ Create distinct work units in the institutes for cyber security administration, engineering, and incident response.
- ✓ Continuously improve employee knowledge of cyber security issues. And so, on

CHAPTER FOUR

TAILORED SOCIAL ENGINEERING ATTACK DETECTION MODEL

4.1 INTRODUCTION

This chapter will establish the rules based on the Data from the Dashen Bank Incident Response Report, the National Cyber security Survey Report (INSA, 2014), and a phishing simulation analysis that will be included in the model. Then, one by one, we will explain each state of the proposed model, and lastly, we will test and validate the model using attack scenarios.

4.2 PSYCHOLOGY BEHIND SOCIAL ENGINEERING AND HUMAN REASONING

The manipulation of humans to acquire illegal access to sensitive information, commit fraudulent operations, or influence behavior for personal advantage is referred to as social engineering. It is a sort of art that uses human psychology and reasoning to fool and influence unsuspecting people. Understanding the psychological concepts underlying social engineering approaches is critical for designing effective counter-strategies. This article delves into the complex interplay between social engineering and human thinking, examining fundamental psychological variables that render people vulnerable to manipulation.

Trust and Authority:

Trust is a key psychological component used in social engineering. Humans have an innate predisposition to trust others, particularly those in positions of authority or apparent competence. Social engineers take advantage of this innate trust by appearing as trustworthy characters such as tech support, bank representatives, or even government authorities. They use this trust to deceive people into disclosing critical information or taking acts that jeopardize their security.

Reciprocity and Obligation:

Reciprocity is a fundamental notion in human relationships in which people feel bound to return benefits or concessions given to them by others. Social engineers frequently start conversations by providing seemingly benign help, favors, or presents. They generate a sense of indebtedness by invoking the reciprocity principle, making people more willing to agree with subsequent demands, even if they appear illogical or questionable.

Consistency and Commitment:

People crave consistency in their views, attitudes, and behaviors. Social engineers take advantage of this cognitive bias by gaining tiny initial commitments or public affirmations from individuals. Individuals are more inclined to continue down a road if they have made a public commitment or taken a small step in that direction, even if following demands or acts are more significant or contradictory to their initial attitude.

Scarcity and Urgency:

According to the scarcity principle, individuals value and seek items that are believed to be scarce or restricted in supply. Social engineers take advantage of this by instilling urgency or scarcity in their demands or offerings. They may use limited-time offers, exclusive access, or impending threats to trick people into behaving quickly and without adequate thought. The increased emotional state caused by haste decreases rational thinking, leaving people more vulnerable to manipulation.

Emotional Manipulation:

Emotions play an important part in decision-making and have the ability to overcome rational judgment. To manipulate others, social engineers use emotional triggers such as fear, excitement, curiosity, or sympathy. They weaken critical thinking and create impulsive reactions by leveraging these emotions, prompting individuals to reveal sensitive information or engage in risky activity that they would otherwise avoid.

Biases in cognition:

Human cognition is prone to biases, which social engineers expertly exploit. Individuals who seek information that validates their pre-existing opinions are subject to deception due to confirmation bias. Social engineers deliver communications that are suited to the target's beliefs, enhancing the possibility of compliance. Other biases, such as authority bias, availability heuristic, and social proof, make people more vulnerable to social engineering attempts.

Therefore, Social engineering thrives on the complexities of human psychology and thinking. Individuals can develop a heightened awareness of social engineering techniques and become more resistant to manipulation by recognizing the underlying psychological variables at work. To educate employees and limit the dangers associated with social engineering attacks, organizations can develop effective training programs and security measures. Finally, education and awareness are crucial in protecting individuals and organizations from social engineering approaches.

4.3 DEFINING RULES

A social engineering attack detection model is a method for detecting and preventing social engineering attacks. To detect suspicious activity or trends, these models often employ a set of rules. The principles are based on understanding of how typical social engineering attacks are carried out.

A social engineering attack detection model's rules can be defined in a variety of ways. A rule-based system is one method. A rule-based system creates a set of rules that describe questionable behavior or patterns.

Based on our data analysis and recommendation from Experts in social engineering, the study developed the following guidelines for inclusion in the proposed social engineering attack detection model:

> **Rule 1:** Be cautious if the requester requests sensitive information such as your personal information or non-public data.
>
> **Rule 2:** Be suspicious if the requester attempts to create a sense of urgency, such as threatening to freeze your account if you neglect the information, and so on.
>
> **Rule 3:** Be wary if the requester urges you to open an attachment or click on a link out of fear or curiosity.

Rule 4: Be cautious if the requester uses a bogus name or email address.

Rule 5: Be alert if the requester threatens you or tries to frighten you.

If any of these red flags appear, it is advised to err on the side of caution and refuse to supply any information to the requester. You should also need to notify your IT department or the authorities about the event.

4.3 TAILORED SOCIAL ENGINEERING ATTACK DETECTION MODEL

The proposed Tailored Social Engineering Attack Detection Model (TSEADM) is primarily concerned with adapting and modifying SEADM version 2 based on study analysis findings, so that it can be used as a tool by individuals to educate themselves about the most recent attack technique and to always be vigilant and on the lookout for social engineering attacks.

This TSEADM employs a decision tree and splits the process into more manageable components to aid decision making as SEADM v2 did. The model also displays the flow of action and how a 'receiver' should handle any form of request. This word refers to the person dealing with the request throughout this discussion, whereas the term 'requester' refers to the person or object who requests the specific action or information from the receiver.

Figure 4.1 below displays the proposed Tailored Social Engineering Attack (TSEADM), which is capable of detecting attacks from all three communication groups.

The model includes a total of 3 types of states: request, receiver, and requester, each of which provides a brief description of what may be expected to happen in each condition. The request states (in yellow) deal directly with information concerning the request itself. The receiver state, shown in blue, deals directly with the person processing the request and whether or not this person (the receiver) understands and is authorized to carry out the request. The requester states, highlighted in green, whether the requester can be directly dealt with and whether any information on the requester can be verified.

This thesis discusses each of these states individually, as indicated in Figure 4.1, before discussing the whole model using examples.

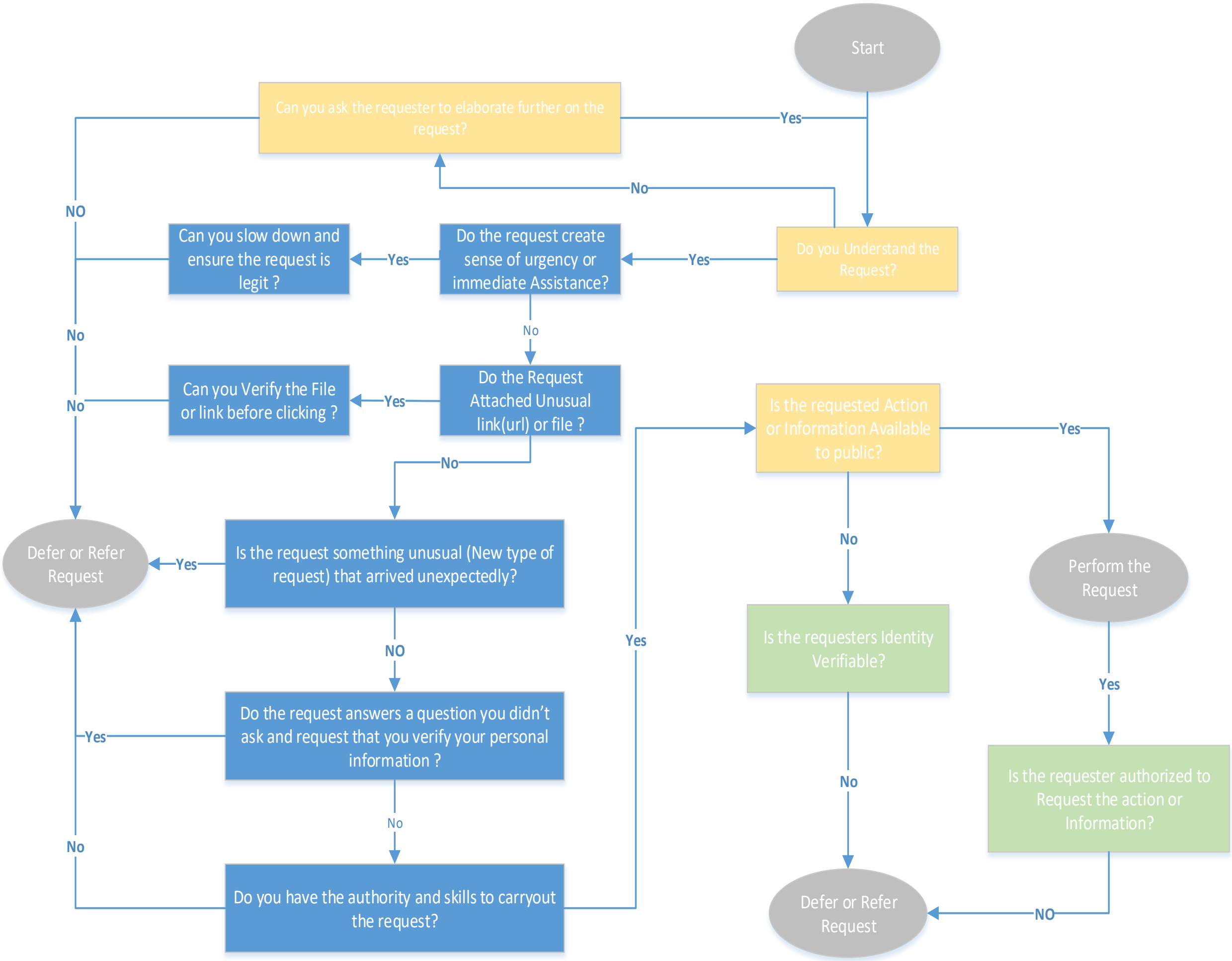

Start
Can you ask the requester to elaborate further on the request?
Yes
No
Do you Understand the Request?
Yes
Do the request create sense of urgency or immediate Assistance?
Yes
Can you slow down and ensure the request is legit ?
NO
No
No
Do the Request Attached Unusual link(url) or file ?
Yes
Can you Verify the File or link before clicking ?
No
Is the request something unusual (New type of request) that arrived unexpectedly?
Yes
Defer or Refer Request
NO
Do the request answers a question you didn't ask and request that you verify your personal information ?
Yes
No
Do you have the authority and skills to carryout the request?
No
Yes
Is the requested Action or Information Available to public?
Yes
Perform the Request
No
Is the requesters Identity Verifiable?
Yes
Is the requester authorized to Request the action or Information?
No
NO
Defer or Refer Request

Fig 4.1 Tailored Social Engineering Attack Detection Model (TSEADM)

Each state in the model has been generalized to the point where it can contain any number of questions necessary to reach a certain transition outcome. This serves as a rough reference for flexibility and extension, depending on the specific environment to which the model is applied.

The states in this model are designed to reach one of the two main states; which are a state to halt the request ($\boldsymbol{S}_F$) or perform the request ($\boldsymbol{S}_S$).

State SF: Halt the Request

This is the negative outcome condition. The request will be halted in this status. In some cases, this may indicate that the request requires further examination or that it is preferable to defer the request to a more authoritative recipient. Deferring the request may also result in the request never being fulfilled and may be seen as the request being halted.

State SS: Perform the Request

This is the model's positive result state. In this state, the receiver is only permitted to carry out the requester's single request.

State S1: Understanding the Request

This state evaluates whether the request's receiver completely understands the request. It means that the requester should have provided all of the information required for the receiver to fully process the request. The TSEADM had the inquiry, "Do you understand what is requested?"

This has been marked as the first state since it is still necessary to completely comprehend the request before it can be handled further.

This state can only transition in one of two ways, as shown below:

- (S_1, U, S_3)
- $(S_1, \neg U, S_2)$

Where;

- **U** denotes that the request has been fully comprehended and that the receiver has all of the information required to carry out the request.
- **¬U** indicates that the request is not fully understood and that the receiver requires additional information.

State S2: Requesting Information to Fully Understand the Request

This state investigates whether the requester can offer enough information for the receiver to fully understand the request. The inquiry in this state was "Can you ask the requester to elaborate further on the request?" in order to determine whether or not the receiver fully understands the request.

This state can only transition in one of two ways, as shown below:

- **(S2, I, S3)**
- **(S2, ¬I, SF)**

Where;

- **I** represent that the requester can and has provided enough information for the receiver to understand the request in its entirety Information required to carry out the request.
- **¬I** represent that the receiver is unable to fully comprehend the request. This could be because the requester was unable to offer additional information, could not be reached, or the information provided by the requester was insufficient or incomplete.

State S3: Does the Requester Attempts to Create a Sense of Urgency?

This state investigates whether the requester attempts to create a sense of urgency or not. this state is associated with the following inquiry to examine if the requester is trying to create a sense of urgency:

- ✓ Does the email use strong language, such as "urgent," "immediately," or "need to act now"?
- ✓ Does the email mention a deadline or time limit?
- ✓ Does the email threaten or imply negative consequences if action is not taken?

This state can only transition in one of two ways, as shown below:

- **(S_3, T, S_4)**
- **(S_3, ¬T, S_5)**

Where;

- **T** represents it's **true** that the requester does create sense of Urgency.
- **¬T** represent it's not true that the requester does create sense of Urgency. Through self-examining the request based on the state questions.

State S4: Can the Receiver Slowdown and Assure the Request Is Legitimate?

This state investigates whether the requester’s sense of urgency is genuine or not through verifying the sender's identity before taking any action. This state is associated with the following inquiry to examine:

- ✓ Is sender's email address from a legitimate company or organization and you recognize the sender?
- ✓ Look for spelling and grammar errors. Which can be a red flag.
- ✓ Does the content of the email ask for personal information, such as your password or credit card number?

This state can only transition in one of two ways, as shown below:

- **(S_4, L, S_5)**
- **(S_4, $\neg L$, S_F)**

Where;

- **L** represents that the receiver can slowdown and assure the request is **legitimate** and need the requested immediate action

- **¬L** represent that the receiver not sure whether an email is legitimate and can’t Slowdown and Assure the Request

State S5: Does Requester Urges You to Open an Attachment or Click on a Link?

This state looks into a typical phishing technique in which the requester attempts to fool you into opening a harmful file or webpage. The attacker can steal your personal information, install malware on your computer, or take control of your computer after you open the file or visit the page.

This state can only transition in one of two ways, as shown below:

- **(S_5, C, S_7)**
- **(S_5, $\neg C$, S_6)**

Where;

- **C** represents that the email did not **contain** any an attachment files or click to open

- **¬C** represent that the email did contain any an attachment files or click to open

State S6: Can the Receiver Verify the Attachment or the Link Before Clicking?

This state looks into ways to verify an attachment or link before clicking on it, this state is associated with the following inquiry to examine:

- ✓ Can you check the sender's email address? If you don't recognize the sender, don't open the attachment or click on the link.
- ✓ Was there any grammatical or spelling errors? Phishing emails often contain errors, which is a sign that they're not from a legitimate source.
- ✓ Can you scan the attachment for viruses? Even if you trust the sender, it's always a good idea to scan attachments for viruses before opening them.
- ✓ Hover your mouse over the link without clicking on it. This will show you the actual URL of the link. Does the URL look suspicious? Don’t click on it.

This state can only transition in one of two ways, as shown below:

- (S_6, V, S_7)
- $(S_5, \neg V, S_F)$

Where;

- **V** represents that the Receiver can Verify the Attachment or the Link Before Clicking
- **¬V** represent that the Receiver cannot Verify the Attachment or the Link Before Clicking and still

State S7: Does the Request Something Unusual and That Arrived Unexpectedly or Answers a Question You Did Not Ask and Request You to Verify Your Personal Information?

This state investigates if the requests something unusual and that arrived unexpectedly or answers a question you did not ask and request you to verify your personal information.

This state can only transition in one of two ways, as shown below:

- (S_7, T, S_F)
- $(S_7, \neg T, S_8)$

Where;

- **T** denotes that it is true. The request is unique and unexpected, or it answers a question you did not ask and asks you to verify your personal information.

- **¬T** indicate that the request is not exceptional and comes unexpectedly, or that it answers a query you did not ask, and that it requests you to verify your personal information.

State S8: Does the Receiver Have the Authority to Perform the Request?

This state investigates whether the receiver has the authority to carry out the request; this state is associated with the following inquiry to examine:

- ✓ Do you have the necessary permissions to complete this request?
- ✓ Can you confirm that you have the authority to carry out this request?
- ✓ Are you able to perform this request on behalf of the company?

If the receiver is unable to confirm that they have the authority to perform the request, you may need to escalate the request to a higher level.

This state can only transition in one of two ways, as shown below:

- $\boldsymbol{(S_8, A, S_9)}$
- $\boldsymbol{(S_8, \neg A, S_F)}$

Where;

- **A** denotes that the receiver is authorized to carry out the request.
- ¬ **A** denotes that the receiver is not authorized to carry out the request.

State S9: Is the Requested Action or Information Available for Public?

This state investigates whether the requested action or information confidential or not. If you are unsure whether the requested action or information is available to the public, it is always best to ask and confirm.

This state can only transition in one of two ways, as shown below:

- $\boldsymbol{(S_9, C, S_{10})}$
- $\boldsymbol{(S_9, \neg C, S_S)}$

Where;

- **A** denotes that the receiver is authorized to carry out the request.
- ¬ **A** denotes that the receiver is not authorized to carry out the request.

State S10: Is the Requester Identity Verifiable?

This state investigates to determine if the requester's identity is verifiable; this state is associated with the following inquiry to examine:

- ✓ Can you verify Requester name and contact information, copy of ID if needed?
- ✓ Can you verify Requester email address or phone number?

If the requester is unable to provide any of this information, you may need to deny their request.

This state can only transition in one of two ways, as shown below:

- $\boldsymbol{(S_{10}, V, S_{11})}$
- $\boldsymbol{(S_{10}, \neg V, S_F)}$

Where;

- **V** denotes that the requester's identity is verifiable.
- ¬ **V** denotes that the requester's identity is verifiable.

State S11: Does the Requester have the Necessary Authority to Request the Action or Information?

This state investigates to determine if the Requester have the Necessary Authority to Request the Action or Information or not, it is important to note that the specific requirements for authority to request information or action may vary depending on the situation

This state can only transition in one of two ways, as shown below:

- $\boldsymbol{(S_{11}, A, S_S)}$

- ***(S_{11}, ¬A, S_F)***

Where;

- **A** denotes that the requester has the Necessary Authority to Request the Action or Information.
- ¬ **A** denotes that the requester lacks the required authority to request the action or information.

4.4 TEST AND VALIDATE THE MODEL

We can put the TSEADM to the test by running it through a variety of social engineering attacks. The model is tested using two sample scenarios supplied in this section (real-world attack examples).

Finally, we assess the model's efficacy based on its ability to detect social engineering attacks reliably and efficiently.

SCENARIO ONE

In this scenario (Pretexting), a social engineer (SE) tries to gather sensitive information about an employee that only they should know. The information is not available to the general public. This attack makes use of bidirectional communication. The social engineer sends an e-mail to a target organization employee. As a result, this e-mail would be sent from a different address than the company's own. To persuade the receiver to perform the specified action of supplying information to the social engineer through the attached link, the social engineer utilizes a pretext and impersonates the responsible HR employee. The remainder of this section connects the example to the model.

Do you understand what is requested?

The e-mail from the social engineer should clearly state what information is required and make the request easily understandable to the receiver. When the receiver understands the request, the ‘yes’ option is selected.

Does the Requester Attempts to Create a Sense of Urgency?

The email does not convey a sense of urgency or instant aid; rather, it invites willing engagement. When the receiver believes there is no sense of urgency in the request, the ‘No’ option is selected.

Does Requester Urges You to Open an Attachment or Click on a Link?

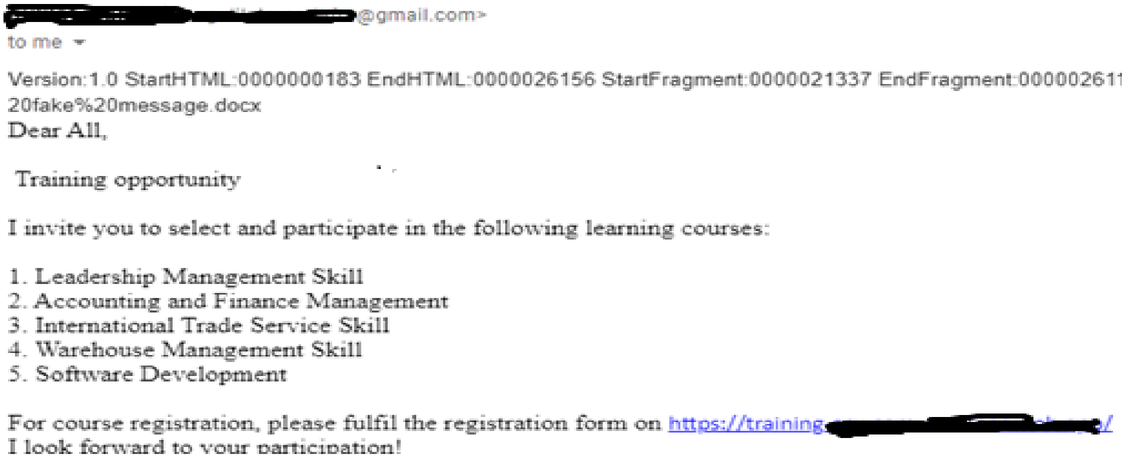
The email wasn't forceful because it invites consensual participation, but it did include a link and asked the receiver to use the link to register for the course described in the pretext. When the receiver believes there is attachment of file or links in the request, the ‘Yes’ option is selected.

Can the Receiver Verify the Attachment or the Link Before Clicking?

In this case the receiver checks the email based on the provided inquiry in this state to verify the file or the link. Which are;

- ✓ Can the receiver Check the sender's email address (is it company email or not)? If you don't recognize the sender, don't open the attachment or click on the link.
- ✓ Was there any grammatical or spelling errors? Phishing emails often contain errors, which is a sign that they're not from a legitimate source.
- ✓ Can the receiver scan the attachment for viruses? Even if you trust the sender, it's always a good idea to scan attachments for viruses before opening them.
- ✓ Hover your mouse over the link without clicking on it. This will show you the actual URL of the link. Does the URL look suspicious? Don’t click on it.

When the receiver is still unsure or one of the inquiry results is affirmative, for example, in this scenario the requester email is not a company email and is simply using the HR employee name to garner trust and obtain the information he sought. As a result, the 'No' option is chosen. When the 'No' option is selected in this state, the model halts the request, either referring it to the security team for additional examination or terminating the request and labeling it as an attack.



[redacted]@gmail.com>
to me

Version:1.0 StartHTML:0000000183 EndHTML:0000026156 StartFragment:0000021337 EndFragment:00000261
20fake%20message.docx
Dear All,

Training opportunity

I invite you to select and participate in the following learning courses:

1. Leadership Management Skill
2. Accounting and Finance Management
3. International Trade Service Skill
4. Warehouse Management Skill
5. Software Development

For course registration, please fulfil the registration form on https://training[redacted]/
I look forward to your participation!

Fig 5.2 Scenario one pretexting attack e-mail

SCENARIO TWO

In this scenario (Phishing mail) The SE identifies a target, an important company employee, and collects information such as their email address and job title. The SE then produces an email that appears to be from a trusted source, in this case the target's IT support team. The email may utilize a domain name that is very similar to the company's genuine domain, but with a minor difference that may go missed at first glance. The email may appear legitimate and professional, complete with the corporate logo and an urgent message requiring rapid action. The email will contain a link that will lead the target to what appears to be the company's login page or another official website.

Do you understand what is requested?

The e-mail from the social engineer should clearly state what information is required and make the request easily understandable to the receiver. When the receiver understands the request, the ‘yes’ option is selected.

Does the Requester Attempts to Create a Sense of Urgency?

The email does create a sense of urgency or instant aid, When the receiver believes there is a sense of urgency in the request, the ‘Yes’ option is selected.

Can the Receiver Slowdown and Assure the Request Is Legitimate?

In this state the Receiver investigates whether the requesters sense of urgency is genuine or not; through verifying the sender's identity before taking any action. This state is associated with the following inquiry to examine:

- ✓ Is sender's email address from a legitimate company or organization and you recognize the sender?
- ✓ Look for spelling and grammar errors. Which can be a red flag.
- ✓ Does the content of the email ask for personal information, such as your password or credit card number?

When the receiver is still unsure or one or more of the inquiry results is affirmative, for example, in this scenario the requester email utilizes a domain name that is very similar to the company's genuine domain, but different and all company email should use same domain name. As a result, the 'No' option is chosen. When the 'No' option is selected in this state, the model halts the request, either referring it to the security team for additional examination or terminating the request and labeling it as an attack.

Therefore, to rate the accuracy/efficiency of this model the study use selected real-world attack examples data set and measures its ability to detect social engineering attacks reliably and efficiently. The study follows the following steps:

1. The study specifies test set data by using 30 high-severity Phishing attacks recorded.
2. Use the TSEADM to make predictions or decisions on the test set based on the predefined rules.
3. Determine the proper or expected outcomes for each of the test examples. Which are all positive in this data set regarding a phishing attack because the test set data are categorized as high-severity phishing attacks in the incident response report.
4. Evaluating whether the model's predictions match the expected outcomes. The result of the evaluation of the test set is summarized in the table below:

Table 4.1 TSEADM test result with real-world attack examples

Rule ID	Rule	Deferred/Referred Request		Request Performed
		Yes	No	
S1	Do you understand what is requested?	Null	30	**NONE**
S2	Can you ask the requester to elaborate further on the request?	Null	30	
S3	Does the Requester Attempts to Create a Sense of Urgency?	13	17	
S4	Can the Receiver Slowdown and Assure the Request Is Legitimate?	13	17	
S5	Does Requester Urges You to Open an Attachment or Click on a Link?	17	Null	
S6	Can the Receiver Verify the Attachment or the Link Before Clicking?	17	Null	
S7	Does the Request Something Unusual and That Arrived Unexpectedly or Answers a Question You Did Not Ask and Request You to Verify Your Personal Information?	Null	Null	
S8	Does the Receiver Have the Authority to Perform the Request?	Null	Null	
S9	Is the Requested Action or Information Available for Public?	Null	Null	
S10	Is the Requester Identity Verifiable?	Null	Null	
S11	Does the Requester have the Necessary Authority to Request the Action or Information?	Null	Null	

Because our test data collection consists of 30 real-world attack instances, the desired outcome of this test should be that all 30 requests are not executed. And the TSEADM prediction reflects this by reducing 13 requests on state 4 and 17 on state 6 and not performing all 30 requests.

Therefore, the accuracy of this TSEADM can be calculated as follow;

- Accuracy: It measures the proportion of correct predictions out of the total predictions made.

Accuracy = (Number of correct predictions) / (Total number of predictions)

= (30) / (30)

Accuracy = ***1***

Accuracy is equals to one means that, the TSEADM always predicts the correct label.

4.5 SUMMARY

The protection of information is critical in modern society, and while information security is always increasing, a weak point remains the human being, who is vulnerable to manipulation techniques.

The Tailored Social Engineering Attack Detection Model (TSEADM) suggested in this chapter is largely focused with adapting and modifying SEADM version 2. The TSEADM has been validated using examples of generalized social engineering attacks.

It was demonstrated that the TSEADM can withstand social engineering attack that leverage both bidirectional and unidirectional communication. The suggested TSEADM can now be utilized to defend against social engineering attempts and can be used as a tool by individuals to educate themselves to be more vigilant and have their guard up always against social engineering attacks.

CHAPTER FIVE

CONCLUSIONS AND FUTURE WORKS

5.1 CONCLUSIONS

Social engineering attacks are on the rise in Ethiopia and around the world, due to a lack of information about the subject, which makes people more vulnerable. There is a need for greater research in this sector to aid in raising awareness regarding social engineering attacks. Awareness and training have been frequently advised as ways for mitigating social engineering attacks because the human being remains a weak point, which social engineers exploit. Having a social engineering attack detection model and educating society about social engineering attacks is required to reduce the growing danger.

As a result, developing a social engineering attack detection model and educating society about social engineering attack is necessary to mitigate the growing threat. This study found that there is a need to expand people's awareness of social engineering, and the recommended approach helps individuals learn about the most recent attack tactic while also being cautious and on the lookout for social engineering attacks.

It was demonstrated that the TSEADM can resist various sorts of communication and tactics used in social engineering attempts. Furthermore, 96% of social engineering assaults enter organizations through email inboxes, according to the 2020 Verizon Data Breach Investigations Report. While most companies rely on individual knowledge to detect social engineering attacks and report on them, having an easy model that incorporates the most recent social engineering attack tactics, such as the one proposed in this thesis paper, is essential to educate individuals and rely on their judgment.

5.2 FUTURE WORKS

The country's cyber security is at stake, according to the data analysis and findings. If this high risk is not mitigated or resolved as quickly as feasible, Ethiopia would face economic, political, and social difficulties.

Therefore, countries' key institutes should do the following to secure their key resource:

- ✓ Implements Critical Mass Cyber Security Standards;
- ✓ Develops And Implements Cyber Security Strategies and Policies;
- ✓ Converts Analog Systems to Digital Systems;
- ✓ Develops and Implements Document Management System.
- ✓ Create Distinct Work Units in The Institutes for Cyber Security Administration, Engineering, And Incident Response.
- ✓ Continuously Improve Employee Knowledge of Cyber Security Issues. And So, Forth

REFERENCES

Francois Mouton (November 2018). “Social Engineering Attack Detection Model.” University of Pretoria.

Fatima Salahdine and Naima Kaabouch. (April 2019). “Social Engineering Attacks: A Survey.” University of North Dako.

Marcel Teixeira. “Literature Review of Social Engineering Detection Models”. Department of Computer Science University of Cape Town.

Andrew M. Colarik. (2016)."The Psychology of Social Engineering: The Science behind Manipulation and Deception”. CRC Press.

Smith, J. A., & Johnson, R. D. (2018). A machine learning approach to social engineering attack detection. Journal of Information Security and Applications, 43, 16-26.

Monique Bezuidenhout and Francois Mouton. (September 2010). “Social Engineering Attack Detection Model (SEADM)”. Information Security for South Africa.

Ibrahim Ghafir, Vaclav Prenosil, Ahmad Alhejailan and Mohammad Hammoudeh. (September 2016) “Social Engineering Attack Strategies and Defense Approaches”. IEEE.

Katharin Krombholz, Heidelinde Hobel, Markus Huber and Edgar Weippl.(June 2015) “Advanced Social Engineering Attacks”. Journal of Information Security and Applications.

Merton Lansleya, Francois Mouton, Stelios Kapetanakis & Nikolaos Polatidis. (March 020). “Seader++: Social Engineering Attack Detection in Online Environments Using Machine Learning”. Journal of Information and Telecommunication.

Jackson, C., & Williams, H. (2019). Detecting social engineering attacks on online social networks using behavioral biometrics. Journal of Computer Security, 27(4), 491-515.

Ansh Mehta, Dev Vora, Harsh Sachala, Jay Khatri and Dharmil Gada. (October 2021). “A Review of Social Engineering Attacks and their Mitigation Solutions”. International Journal of Engineering Research & Technology (IJERT).

Robert W. Taylor. (2014). “Social Engineering: An Insider Threat". Information Security Journal: A Global Perspective.

Emma L. Williams. (2017). “Social Engineering: The Role of Deception in Cyber Security". Journal of Strategic Security.

Brown, C., & Jones, K. (2020). Detecting social engineering attacks using an ensemble of machine learning classifiers. In Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES) (pp. 1-10). IEEE.

John R. Vacca. (2016). "Social Engineering: Manipulation and Deception in Cyber Attacks”. Elsevier.

Lee, S., Kim, Y., & Park, J. (2021). An advanced social engineering attack detection model using deep learning algorithms. Electronics, 10(3), 283.

Patel, D., & Joshi, P. (2019). Social engineering attack detection using machine learning and cognitive approach. In Proceedings of the International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM) (pp. 15-24). Springer.

Richard D. Kuehn and Katherine E. Heaton. (2012). "A Framework for Understanding and Applying Social Engineering Principles" Journal of Information Warfare.

Kevin Mitnick and William L. Simon. (2017). “Social Engineering: The Art of Human Hacking, 2nd Edition”. John Wiley & Sons.

Christopher Hadnagy. (2009). "Social Engineering: Understanding and Combating Threats to Information Security". Security Journal.

James Messer. (2012). “Social Engineering: A Comprehensive Guide to the Human Factor in Security”.

Bruce Schneier. (2011). “Social Engineering Attacks: How to Protect Yourself “. SANS Institute.

K. Mathiyalagan and R. Anand. (2012). “Social Engineering: A Threat to Information Security". International Journal of Computer Applications.

INSA. (2014). “**ሀገራዊ የሳይበር ደህንነት ዳሰሳ ሪፖርት**”. https://www.insa.gov.et/web/en/documents.