SOCIAL ENGINEERING ATTACK DETECTION MODEL

by

Francois Mouton

Submitted in fulfillment of the requirements for the degree
Philosophiae Doctor (Computer Science)

in the

Department of Computer Science
Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

November 2018

SUMMARY

SOCIAL ENGINEERING ATTACK DETECTION MODEL

by

Francois Mouton

Promoter(s):	Prof H.S. Venter
Department:	Computer Science
University:	University of Pretoria
Degree:	Philosophiae Doctor (Computer Science)
Keywords:	Bidirectional Communication, Computer Security, Deontology, Indirect Communication, Research Ethics, Social Engineering, Social Engineering Attack Detection Model, Social Engineering Attack Examples, Social Engineering Attack Framework, Unidirectional Communication, Utilitarianism, Virtue Ethics.

Information security is a fast-growing discipline, and relies on continued improvement of security measures to protect sensitive information. Human operators are one of the weakest links in the security chain as they are highly susceptible to manipulation. A social engineering attack targets this weakness by using various manipulation techniques to elicit individuals to perform sensitive requests. Social engineering is deeply entrenched in the fields of both computer science and social psychology. Knowledge is required in both these disciplines to perform social engineering based research.

The field of social engineering is still lacking with regards to standardised definitions, ethical concerns, attack frameworks, examples of attacks and detection models. The main focus of this thesis is the proposal of a social engineering attack detection model, however, this thesis also addresses gaps within the field with regards to standardised definitions, ethical concerns, attack frameworks and examples of attacks.

The first step of this journey was to review the existing definitions within the field of social engineering. After the review, this thesis proposed standardised definitions for *social engineer*, *social engineering*, *social engineered* and *social engineering attack*. It was also established that social engineering can only be performed over bidirectional, unidirectional and indirect communication.

This thesis also identifies a number of concerns regarding social engineering in public communication, penetration testing and social engineering research. It also discusses the identified concerns with regard to three different normative ethics approaches (virtue ethics, utilitarianism and deontology) and provides their corresponding ethical perspectives.

Furthermore, this thesis proposes a social engineering attack framework based on Kevin Mitnick's social engineering attack cycle. The attack framework addresses shortcomings of Mitnick's social engineering attack cycle and focuses on every step of the social engineering attack from determining the goal of an attack up to the successful conclusion of the attack.

The social engineering attack framework is then utilised to derive detailed social engineering attack examples from real-world social engineering attacks within literature. Mapping several similar real-world examples to the social engineering attack framework allows one to establish a detailed flow of the attack whilst abstracting subjects and objects. This mapping is then utilised to propose the generalised social engineering attack examples that are representative of real-world examples, whilst still being general enough to encompass several different real-world examples.

After all of the gaps within the field of social engineering were addressed, attention is shifted back towards the main focus of this thesis which is the social engineering attack detection model. There were three iterations of the social engineering attack detection model proposed throughout this thesis, with each iteration improving upon the limitations on the one prior. The first iteration of the social engineering attack detection model was designed with a call centre environment in mind and is only able to cater for social engineering attacks that use bidirectional communication. The second iteration of the social engineering attack detection model addresses this problem by extending the model to cater for social engineering attacks that use either bidirectional communication, unidirectional communication or indirect communication. The third iteration focuses on the proposal of the underlying finite state machine of the social engineering attack detection model. The third iteration of the social engineering attack detection model provides a more abstract and extensible model that highlights the

inter-connections between task categories associated with different scenarios. Furthermore, the third iteration is intended to help facilitate the incorporation of organisation specific extensions by grouping similar activities into distinct categories, subdivided into one or more states. In addition, it facilitates additional analysis on state transitions that are difficult to extract from the second iteration.

Ultimately, this thesis proposes a refined social engineering attack detection model that can be utilised by industry to either implement into their environment or to be used as a social engineering awareness training tool. The social engineering attack detection model is also developed to be extensible so that other researchers can expand upon the proposed model. Lastly, the social engineering attack detection model can also be used as a comparative measure for future social engineering attack detection models.

Acknowledgements

I would like to express my sincere thanks to the following people for their assistance during the production of this dissertation:

- **My parents, Kobus and Jane Mouton** — Thank you for your loving support both throughout my studies and in all the decisions I have made during my life.

- **Hein Venter** — Thank you for your patience, inspiration and motivation. Most of all, thank you for your friendship while I was under your supervision.

- **Work Colleagues** — Thank you for your assistance with proof reading whenever I required the assistance.

- **Members of the HAISA Conference** — Thank you to the great friendships that were established through many years of this conference.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

Information security is a fast-growing discipline. The protection of information is of vital importance to organisations and governments, and the development of measures to counter illegal access to information is an area that receives increasing attention. Organisations and governments have a vested interest in securing sensitive information and hence in securing the trust of clients and citizens. Technology on its own is not a sufficient safeguard against information theft; staff members are often the weak link in an information security system. Staff members can be influenced to divulge sensitive information or to perform an unauthorised task, which subsequently allows unauthorised individuals access to protected systems. This notion of influencing humans to divulge sensitive information or to perform an unauthorised task has been coined as the term 'social engineering'.

The remainder of Chapter 1 provides the motivation for this research, gives the reader a brief overview of social engineering and discusses the foreseeable problems within the field. In the next section the reader is introduced to the problem on which the researcher has focused and is given a brief overview of how the problem will be addressed. Chapter 1 concludes with a layout section of the chapters to follow.

1.1 MOTIVATION AND RESEARCH GAP

The term 'social engineering' only started appearing in literature from 1987 onwards. The earliest literature that the author found on social engineering (SE) is an article by Quann and Belford (1987) [1]. According to these authors SE, whilst still in its infancy, is seen as "an attempt to exploit the help desks and other related support services normally associated with computer systems" [1]. SE was later described as "trickery and deceit, also known as Social Engineering", according to Kluepfel (1989) [2, 3]. During the 1990s, the term was rarely used and it only became a household term around the

early 2000s after Kevin Mitnick published two books entitled “The Art of Deception” and “The Art of Intrusion” [4, 5].

This popularisation of the term social engineering also came with some unplanned consequences. There was no commonly agreed upon definition for what social engineering really is and there are mixed opinions on what really forms part of social engineering. As an example, a group of security enthusiasts were asked “What is social engineering?” and their responses varied considerably [6]:

- “Social engineering is lying to people to get information.”
- “Social engineering is being a good actor.”
- “Social engineering is knowing how to get stuff for free.”

All of these answers encompass only a small part, which all forms part of what social engineering entails.

During the same time social engineering became an important topic, communication via e-mail became more widespread because it became the preferred medium for corporate communication over traditional mediums such as postal letters and faxes. The cost and effort of sending several thousand messages are significantly less than sending thousands of postal letters or faxes and this leads to an abundance of scams, especially 419 scams [7, 8]. The name 419 scam originates from the Nigerian Criminal Code which outlaws the practice of the scam. The 419 scam is basically an e-mail, postal letter or any communication in which a target is informed that he or she has been selected for a lucrative opportunity. The target is then requested to provide the sender (attacker) with some assistance, usually financial, in order to benefit from the opportunity [7, 8]. These 419 scams form part of a field called phishing. However, since the ultimate goal of the 419 scam or phishing e-mail is to convince a target to comply with the requested action, it is also deemed to be a social engineering attack. Due to phishing being much better defined than social engineering at the current stage, a significant amount of the research performed within the field of social engineering is specifically focused on phishing attacks.

The instant popularisation of the field, with no standardised definition, as well as having a significant amount of research within the field of social engineering, primarily focusing on the phishing aspect,

have had a significant impact on the field. The biggest impact the author has noticed, is that the research within the field of social engineering does not follow any standardised definition. Each researcher defines social engineering to be specific to their task, which then usually include only an expansion of what phishing encompasses. In the author's opinion, having no formalised definition for social engineering has had several secondary consequences. The field of social engineering does not have any formalised way to determine what it actually entails due to the lack of a formal definition. As a result of this, there is also no standardised attack framework that details the full process of a social engineering attack. Having a standardised attack framework would have helped to ensure that one can correctly and comprehensively document a social engineering attack. Since there is no existing attack framework, there are limited social engineering attack examples documented. Having so little research on how to perform a social engineering attack has also caused the detection side of social engineering to be ignored.

This introduction was aimed at informing the reader of the author's motivation for the research and the research gap within the field of social engineering. The next section explores the problem statement that this study focused on.

1.2 PROBLEM STATEMENT

The overarching problem that this thesis addresses is that there is currently no formalised method or model for individuals to utilise and especially for educating themselves to be more vigilant against social engineering attacks. To solve this comprehensive problem, it is necessary to divide it into smaller problems.

The first aspect that needs to be addressed is the fact that there is currently no standardised definition for the prominent terms within the field of social engineering, namely social engineer, social engineering, social engineered and social engineering attack. This problem is addressed by performing a literature survey of the currently available work within the field of social engineering and using this information to propose formalised definitions for each of the prominent terms within the field.

Solving the first problem allows one to further delve into the field to examine social engineering attacks. However, before social engineering attacks can be researched, the ethical considerations of performing

social engineering attacks and social engineering research should first be considered. The second problem to address is the fact that there is currently little or no research studies available on the ethical considerations that should be taken into account whilst performing either social engineering attacks or social engineering research. This thesis addresses this problem by applying the three mainstream normative ethical perspectives, namely Deontology, Utilitarianism and Virtue Ethics to attack examples which fit the standardised definition of social engineering.

Having the ethical considerations which need to be taken into account whilst performing social engineering research, this thesis is able to address the rest of the short-comings within the field of social engineering. The third problem to address is that there is currently no model that depicts all the phases and steps of a social engineering attack, according to a standardised definition of social engineering attack, onto which already performed or future planned social engineering attacks can be mapped. This problem is addressed by proposing a social engineering attack framework which depicts all of the phases and steps which needs to be performed during a social engineering attack.

Having the proposed social engineering attack framework, allows one to map documented social engineering attacks. During this process of mapping documented social engineering attacks onto the framework, the fourth problem was discovered. The fourth problem to address is that there are limited examples documented of social engineering attacks and most of the documented examples do not include all of the information required to populate all of the phases and steps within the social engineering attack framework. Also, due to the need to be able to test and verify the social engineering attack detection model and the ethical considerations of performing social engineering attacks, this thesis proposes standardised social engineering attack examples. The proposed social engineering attack examples contain all of the information to populate each phase and step within the social engineering attack framework. Having these complete social engineering attack examples, allows one to test and verify social engineering models without the need to perform social engineering attacks on individuals and thus allowing one to test and verify social engineering models without violating any ethical constraints of social engineering research.

After addressing all four of the problems described above, the overarching problem of this thesis could be addressed, namely the lack of a formalised method or model which individuals can utilise and educate themselves with in order to be more vigilant against social engineering attacks. To address this problem, the author studied previous research on social engineering attack detection models. All of

the above-mentioned research are taken into account to propose a social engineering attack detection model which will be able to cater for all forms of social engineering attacks that forms part of the standardised definition of social engineering attacks. This model is then further verified using the proposed social engineering attack examples in order to verify the completeness and correctness of the social engineering attack detection model.

The following section provides more detail on the goals that have been identified in order to solve the problems mentioned above and the objectives that this research aims to fulfil.

1.3 RESEARCH GOALS AND OBJECTIVES

The primary goal of the study is to develop and propose a Social Engineering Attack Detection Model (SEADM). The author has several research objectives in mind that is directly associated with the primary goal:

- The SEADM should be able to cater for various types of social engineering attacks and have no specific focus on a singular type of attack.
- The SEADM should provide a step by step process of identifying social engineering attacks.
- The use of the SEADM must facilitate education and training for users to be more vigilant against social engineering attacks. that will also educate and train users to be more vigilant against such attacks after each request is put through the model.
- The SEADM should be simplistic so that it can be used by any individual without requiring specific training on the model itself.

Additional to developing the SEADM, the secondary goal of this study is to substantially add to and grow the field of social engineering. In order to achieve this goal, the following objectives needs to be pursued:

- The field of social engineering must be enhanced by exploring and proposing standardised definitions.

- Ethical considerations for research in the social engineering domain must be investigated.

- Social engineering attack frameworks must to be investigated.

- Social engineering attack examples must be explored.

All of these topics within the field of social engineering are currently not fully explored and needed to be expanded upon in order to perform the primary goal of this study.

The following section provides the reader with an overview of the contribution of this study within the field of social engineering.

1.4 RESEARCH CONTRIBUTIONS

This study contibutes the following definitions, frameworks, concepts and models to the field of social engineering:

- Standardised definitions for the terms social engineer, social engineering, social engineered and social engineering attack.

- The ethical concerns regarding social engineering from the ethical perspectives of Deontology, Utilitarianism and Virtue Ethics.

- Proposing a social engineering attack framework which caters for all types of social engineering attacks as encompassed by the standardised definition of social engineering attack.

- Proposing social engineering attack examples, applied to the social engineering attack framework, which can be used to test and verify the completeness and performance of models within the field of social engineering without requiring to perform the social engineering attacks.

- Proposing a social engineering attack detection model which can be utilised as a tool by individuals to educate themselves to be more vigilant against social engineering attacks.

The following section discusses the layout of the dissertation and provides a brief preview of each of the chapters that follow.

1.5 OVERVIEW OF STUDY

Chapter 1 contains the introduction wherein the motivation for the research, the problem statement, the goals and objectives and the contributions of this study are provided. Each of the subsequent chapters contains an introductory chapter motivation. These chapter motivations serves to provide the reader with the rationale for inclusion of the chapter and the logical sequence within the overarching study.

Chapter 2 provides the reader with literature on social engineering and a brief overview of existing models within the field. The literature is intended to familiarise the reader with the current state of the field of social engineering and to focus on where the research gaps currently exist. Due to the importance of having a standardised definition for social engineering throughout this study, the standardised definitions are also provided in this chapter. In this manner, this chapter therefore already contains some contribution work. Furthermore, this chapter proposes both a social engineering attack classification as well as a social engineering attack ontological model. All these elements are used throughout the thesis and therefore it is important to familiarise the reader with these concepts in this early chapter.

Chapter 3 focuses briefly on the importance of SEADMs and why humans are instinctively prone to fall prey to social engineering attacks. This chapter also contains the initial iteration of the SEADM and a discussion on why it was not adequate for detecting all forms of social engineering attacks.

Chapter 4 provides the reader with the ethical considerations which had to be upheld whilst performing this study. This chapter identifies a number of ethical concerns regarding social engineering in public communication, penetration testing and social engineering research. It also discusses the identified ethical concerns with regard to three different normative ethics approaches (Virtue Ethics, Utilitarianism

and Deontology) and provides their corresponding ethical perspectives as well as practical examples of where these formalised ethical concerns for social engineering research can be beneficial.

Chapter 5 proposes a social engineering attack framework based on Kevin Mitnick's social engineering attack cycle. The attack framework addresses shortcomings of Mitnick's social engineering attack cycle and focuses on every step of the social engineering attack from determining the goal of an attack up to the successful conclusion of the attack. The ontological model is used as the basis for the social engineering attack framework, however, the social engineering attack framework expands on the ontological model by being able to additionally represent temporal data such as flow and time.

Chapter 6 proposes detailed social engineering attack examples based on the social engineering attack framework. The proposed social engineering attack examples attempt to alleviate the problem of limited documented literature on social engineering attacks. Current documented examples of social engineering attacks do not include all the attack phases. In order to perform comparative studies of different social engineering models, processes and frameworks, it is necessary to have a formalised set of social engineering attack examples that are fully detailed in every phase and step of the process. The proposed social engineering attack examples cover all three different types of communication, namely bidirectional communication, unidirectional communication and indirect communication. These examples can be used by other researchers to either expand on, use for comparative measures, create additional examples or evaluate models for completeness. These examples are used in this study to be able to determine the completeness and effectiveness of the proposed SEADM.

Chapter 7 provides the reader with the second iteration of the SEADM. The initial iteration of the model, as discussed in Chapter 3, were designed with a call centre environment in mind and is only able to cater for social engineering attacks that use bidirectional communication. Based on the research performed in Chapter 2 it was discovered that social engineering attacks can be classified into three different categories, namely attacks that utilise bidirectional communication, unidirectional communication or indirect communication. The proposed SEADM addresses this problem by extending the model to cater for social engineering attacks that use bidirectional communication, unidirectional communication or indirect communication. This chapter also maps the social engineering attack examples to the SEADM. The social engineering attack examples are used to verify both the completeness and effectiveness of the SEADM. Using the social engineering attack examples to verify the SEADM, one is able to simulate full social engineering attacks, within all three different types

of communication, namely bidirectional communication, unidirectional communication and indirect communication. Furthermore, one is able to verify the SEADM in an ethical manner as the simulations can be executed without having to perform the attacks on individuals as the social engineering attack examples already contain all the phases and steps of a social engineering attack.

Chapter 8 provides the reader with the third and final iteration of the SEADM. This chapter aims to improve the extensibility of the SEADM, and to reduce its implementation complexity by restructuring the process to be cycle-free and deterministic. The SEADM has already been shown to successfully thwart social engineering attacks utilising either bidirectional communication, unidirectional communication or indirect communication. Proposing and exploring the underlying finite state machine of the model allows one to have a clearer overview of the mental processing performed within the model. While the current model provides a general procedural template for implementing detection mechanisms for social engineering attacks, the finite state machine provides a more abstract and extensible model that highlights the inter-connections between task categories associated with different scenarios. The finite state machine is intended to help facilitate the incorporation of organisation specific extensions by grouping similar activities into distinct categories, subdivided into one or more states.

Chapter 9 performs a critical evaluation on all the contributions related to this thesis. This chapter takes the reader on a swift journey through all of the research contributions. Each of the contributions is individually revisited and both the advantages and the disadvantages on each of the contributions are discussed. The chapter concludes with a summary table of the research contributions with their associated advantages and disadvantages.

Chapter 10 concludes this study with a brief summary of contributions of this thesis. This chapter aims to revisit both the problem statement and the research goals for this study. Each of the research contributions, as listed in section 1.4, is also revisited and a discussion is provided to what extent the research contributions address both the problem statement and research goals. This chapter is concluded by proposing areas of future research.

Chapter 1 is now concluded with a graphic representation of the layout of the dissertation as depicted in Figure 1.1.

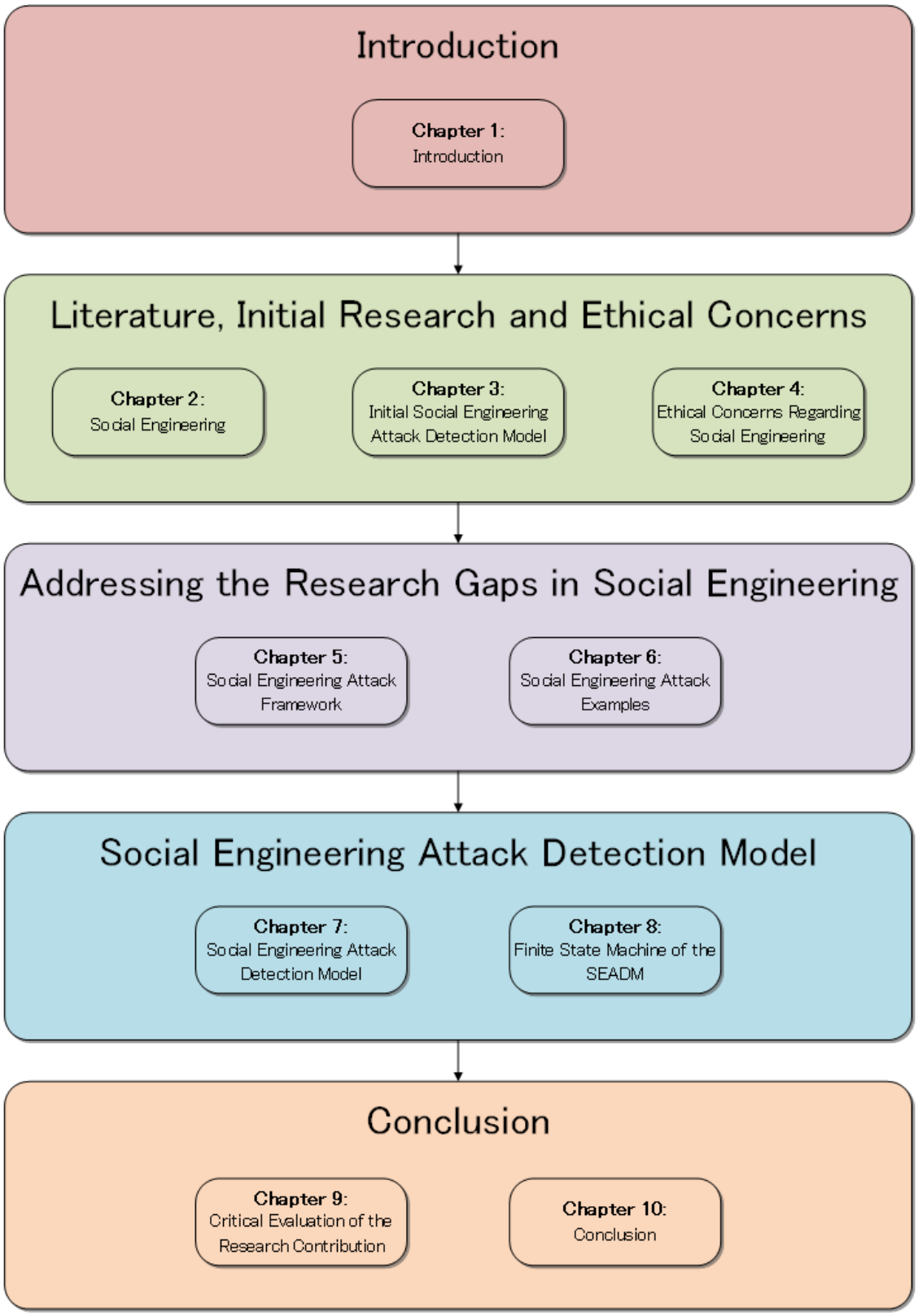

Figure 1.1: Graphical Representation of Dissertation Layout

CHAPTER 2 SOCIAL ENGINEERING

2.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the required literature on social engineering and the historical origin for social engineering. This chapter is based on a conference paper entitled "Towards an Ontological Model Defining the Social Engineering Domain" which has previously been published by the author [9]. This paper was peer reviewed and published at the 11th IFIP International Conference on Human Choice and Computers in 2014 [9]. This conference paper was the first work within the field which examined all of the existing definitions of social engineering in an attempt to develop a consolidated definition. At the start of this thesis, there was no agreed upon definition for social engineering, thus making the need for a consolidated definition vitaly important. It was identified during the course of this research that there is limited literature on social engineering as the field is still in its infancy. The author is aware that this section is not limited to a literature review, as only the historical overview, the overview of existing definitions and attack frameworks are existing research. This chapter already provides contributions by proposing consolidated definitions for *social engineer*, *social engineering* and *social engineering attack*. The motivation for the early inclusion of this chapter lies in the need for the reader to have access to the proposed definitions early on, as it forms the basis of this thesis.

2.2 INTRODUCTION

Social Engineering (SE) is focused on the exploitation of a human in order to gain unauthorised access to information and falls under the umbrella of the Information Security spectrum [10, 11]. Humans are the focus point of most organisations, but they also pose a risk to the sensitive information of the

organisations. Examples of sensitive information are the secret recipe that gives the Kentucky Fried Chicken meals their distinctive flavour, and the personal banking information of clients.

Employees can also be a threat to an organisation due to the insider threat attack vector [12]. Social engineering differs from the insider threat attack vector, as the target of the SE is the human, whereas an insider threat is the employee themselves posing a risk to the organisation.

Although organisations usually employ advanced technical security measures to minimise opportunities for unauthorised individuals to gain access to sensitive information, it is vital that they consider the risk of their staff members falling victim to SE attacks. Humans often react emotionally and thus may be more vulnerable than machines. The biggest threat to an organisation with sensitive information is not the technical protection, but the people who form the core of the organisation. Attackers have realised that it is easier to gain unauthorised access to the information and communications technology infrastructure of an organisation through an individual, rather than trying to penetrate a security system.

Social engineering was not always seen as a field within Information Security. The term social engineering originated from the field of social science. The term "Social engineer" was first found in a news article from the New York Times [13]. The term was used to indicate that Mr T. Burnett Baldwin, as an able social engineer, ensured that the carnival programme has been executed in full detail by the people who he had to manage [13]. From this it is clear that the term was used to depict a person who was able to influence a group of individuals to perform a specific task to the best of their ability. In 1899, the term "Social engineering" was first coined by Mr William Tolman who proclaimed social engineering to be one of the newest professions [14]. In the article, it is explained how Mr Tolman suggested that an organisation converts an empty lot, opposite of the employees' workplace, to a recreation ground and resting place for their wives and children. Consequently, this has led to a much better relationship between the employees and the employer. This is another clear example of how social engineering refers to exerting influence over a group of individuals in an effort to strengthen their relationship with a specific entity.

During the 1900's, the term social engineering was used quite a lot during times of war [15]. Social engineering was tool to promote propaganda either for or against ideologies such as communism, capitalism or socialism [15]. Social engineering is also used with regards to political influences. In

1992 an article was published in the New York Times with the headline "Bush Paints Clinton as 'Social engineer"' [16]. This article depicts social engineering being used for malicious purposes as, according to the article, Bill Clinton is misleading people on the impact on their daily lives with the proposed changes to tax legislation. Even though, it has been shown that social engineering has some history in the field of social sciences, this thesis focuses on social engineering in computer science.

Social Engineering still constitute a relatively new area of research in computer science and the first papers on SE were only published during the late 1980's [1, 2]. The earliest literature that the author found on SE is an article by Quann and Belford (1987) [1]. According to these authors, SE, whilst still in its infancy, is seen as "an attempt to exploit the help desks and other related support services normally associated with computer systems" [1]. SE was later described as "trickery and deceit, also known as Social Engineering", according to Kluepfel (1989) [2, 3]. Even in one of the most prominent hacker magazines, the 2600: The Hacker Quarterly[1], the term *Social Engineering* was not widely used. One of the articles titled, "Janitor Privileges", explains in great detail how to perform an SE attack, however the term *Social Engineering* is never mentioned in the article [18].

Due to the infancy of social engineering within the computer science domain, there is limited research within the field and no globally accepted definition for the field. The purpose of this chapter is to provide a background to social engineering whilst also proposing formalised definitions that will be used throughout this thesis. This chapter also proposes an ontological model for social engineering attacks in order to further the support the formalised definition. Several papers, each with a different view on SE, have been studied and analysed to achieve a singular, all encompassing definition for social engineering.

The remainder of the chapter is constructed as follows. Section 2.3 provides a background on different existing SE definitions and proposes more structured definitions for terms within the domain of SE. Section 2.4 discusses several existing taxonomies for the SE domain. Section 2.5 expands on the definitions provided in section 2.3 by providing an SE attack classification model. Section 2.6 further supports the definitions by providing an ontological model for SE attacks. Section 2.7 concludes this chapter by providing a summary of the contributions.

[1] A magazine which was established by Emmanuel Goldstein in mid January 1984 and contains articles regarding the undergound world of hacking. The individuals publishing in this magazine are mostly individuals who are already facing several charges regarding computer related crimes. [17]

2.3 DEFINING SOCIAL ENGINEERING

In a 1995 publication, the authors Winkler and Dealy posit that the hacker community has started to define SE as "the process of using social interactions to obtain information about a victim's computer system." [19]. The most popular definition of SE is the one by Kevin Mitnick who defines it as "using influence and persuasion to deceive people and take advantage of their misplaced trust in order to obtain insider information" [4].

Various articles define SE and give descriptions of an SE attack. The definitions are diverse and often reflect one aspect of an approach relevant to a particular research project. Commonly agreed upon definitions that include all the different entities in SE are required. The author has identified a large sample of definitions from literature and it has become abundantly clear that no single formalised definition exists.

In order to overcome this, and to have a formalised definition for social engineering, the author has studied and analysed several papers, each with a different view on SE, to achieve a singular, all encompassing definition for SE. The following definitions of SE illustrate that there exists no single, widely accepted definition:

- "a social/psychological process by which an individual can gain information from an individual about a targeted organization." [20]
- "a type of attack against the human element during which the assailant induces the victim to release information or perform actions they should not." [21]
- "the use of social disguises, cultural ploys, and psychological tricks to get computer users to assist hackers in their illegal intrusion or use of computer systems and networks." [22, 23]
- "the art of gaining access to secure objects by exploiting human psychology, rather than using hacking techniques." [24, 25]
- "an attack in which an attacker uses human interaction to obtain or compromise information about an organization or its computer system." [26, 27, 28, 29]

- "a process in which an attacker attempts to acquire information about your network and system by social means." [30, 31]

- "a deception technique utilized by hackers to derive information or data about a particular system or operation." [32, 33, 34]

- "a non-technical kind of intrusion that relies heavily on human interaction and often involves tricking other people to break normal security procedures." [35]

- "a hacker's manipulation of the human tendency to trust other people in order to obtain information that will allow unauthorized access to systems." [36, 37]

- "the science of skilfully manoeuvring human beings to take action in some aspect of their lives." [38, 6]

- "Social Engineering, in the context of information security, is understood to mean the art of manipulating people into performing actions or divulging confidential information." [39]

- "the act of manipulating a person or persons into performing some action." [40, 41]

- "using subversive tactics to elicit information from end users for ulterior motives." [42]

- "using influence and persuasion to deceive people and take advantage of their misplaced trust in order to obtain insider information." [4, 43, 44, 45, 46]

- "the use of social disguises, cultural ploys, and psychological tricks to get computer users to assist hackers in their illegal intrusion or use of computer systems and networks." [47]

These definitions specify different ideas as to what SE involves. Two of these definitions specifically focus on gaining information from an organisation [26, 20, 27, 28, 29]. Several of the definitions sketches SE as the manipulation and persuasion of people in order to get information or to persuade someone to perform some action. Furthermore, some of the definitions are formed around gaining access to computer systems and networks. The only element that all of these definitions have in

common is that a human is exploited in order to gain some unauthorised information or perform some action. As one can see from the vast array of these definitions, there is no single harmonised definition.

The author of this thesis proposes the following harmonised 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 organisation to comply with a specific request from an attacker where either 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* (verb[2]): 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.

These definitions are provided here to equip the reader with the same understanding of the terms as the author. Section 2.6 elaborates more on these definitions. It is also important to fully understand the concept and requirements of a social engineering attack. Apart from the several definitions available for SE, there are also various taxonomies which try to encapsulate SE and the structure of an SE attack. In existing literature, a few taxonomies were proposed to provide some structure to the domain of SE. All of these taxonomies focuses on different elements of social engineering and the taxonomies are discussed in the following section.

[2]The term social engineer in a verb form, can only be written in the Past Perfect form.

2.4 EXISTING SOCIAL ENGINEERING TAXONOMIES

Several taxonomies are studied and discussed in this section: Harley [48], Laribee [49], Ivaturi & Janczewski [50], Mohd et al. [51] and Tetri & Vuorinen [52]. The selected taxonomies were the most cited articles at time of writing and they were all cited between 15 to 35 times each.

2.4.1 Harley (1998)

Harley [48] is one of the first articles to present a taxonomy for the SE domain, and in fact proposes two different taxonomies. The first one defines the following SE techniques and related attacks: Masquerading, Password stealing, Dumpster diving, Leftover, Hoax Virus Alerts and other Chain Letters, Spam and Direct Psychological Manipulation. This taxonomy mixes social compliance principles with techniques.

The second taxonomy defines seven user vulnerabilities: Gullibility, Curiosity, Courtesy, Greed, Diffidence, Thoughtlessness and Apathy. Even though these vulnerabilities are mostly the reasons why individuals are susceptible to SE attacks, they do not specify an SE attack as such. The same SE attack can be performed using more than one of the mentioned vulnerabilities, which clarifies that these vulnerabilities are not the unique establishment of what an SE attack entails. Vulnerabilities of a human, not limited by the seven mentioned above, lead to the susceptibility of an attack.

2.4.2 Laribee (2006)

Laribee [49] identifies two different models, namely a trust model and an attack model. According to Laribee, SE is complex and typically requires multiple communications and targets. The two models are meant to be applied, individually or together, at various times to attain each individual attack goal [49]. The trust model describes how the social engineer establishes a trustworthy relationship with the target, whilst the attack model describes how a social engineer performs an information gathering attack. The attack model is limited to four techniques: deception, manipulation, persuasion and influence. In the attack model the social engineer is only able to use one of these techniques. Furthermore, after the technique has been performed, the attack model feeds into the trust model where the aim is to build a trustworthy relationship. These models are problematic because not all

SE attacks require a continuous relationship since there is not always the need to build a trustworthy relationship with the target. A social engineer generally uses a combination of compliance principles and techniques to perform a single SE attack.

2.4.3 Ivaturi & Janczewski (2011)

Ivaturi & Janczewski [50] classify an SE attack to be either *person-person* or to be *person-person via media*. *Person-person* is when there is direct communication involving a human, whereas *person-person via media* involves some medium used to communicate. The medium can be text, voice or video. Person-person attacks involve impersonation. Different techniques are described.

This taxonomy contains a well-defined structure for different SE techniques, as well as the types of attacks in which they are used. It is very similar to the structure of the direct communication part of the author's model, as further on proposed in section 2.5. Their study only focuses on direct communication and does not further elaborate on a scenario where indirect communication can be used for an SE attack.

2.4.4 Mohd et al. (2011)

Mohd et al. [51] classify an SE attack as being either human-based or technical-based. Human-based attacks apply some techniques that are combined to form an attack, such as "in person" and "simple persuasion". The one technique cannot be used without the other. The items regarded as types of attacks are techniques that form a single attack, rather than being used separately as individual attacks.

Their technical-based attacks are mediums used within an SE attack such as "Email, Software, Web sites". Another example is "Denial of Service" which is not an SE attack; it is an attack on a service and brings down a system instead of extracting information from it. The latter effect is the aim of an SE attack.

In summary, the Modh et al. model describes techniques used in SE attacks instead of depicting an SE attack as a whole.

2.4.5 Tetri & Vuorinen (2013)

Tetri & Vuorinen [52] studied several papers on SE and critically analysed them in order to present an overview of SE. They defined three main dimensions of SE: persuasion, fabrication and data gathering.

Persuasion involves getting someone to comply with an inappropriate request. The paper identifies two features of persuasion: *Direct interaction* and *active engagement between the intruder and the target* [52]. Fabrication involves techniques such as impersonation and using a false identification document to deceive victims into thinking the attacker is someone else. Data gathering is the process of gaining information from the target.

The author of this thesis agrees with Tetri & Vuorinen's description of persuasion although it can be seen as a compliance principle from a psychological perspective. The definitions of fabrication and data gathering on the other hand, are techniques aimed at aiding an SE attack rather than being a phase of an SE attack.

This chapter has now provided the existing research within the field of social engineering with regards to the definition of the field. As part of this thesis, the overarching purpose is to expand on the field and to formalise it further, thus the rest of this chapter is dedicated to use the background information to show how it supports the proposed standardised definitions. The author takes these taxonomies into account and attempt to improve on these ideas by identifying three different subcategories of an SE attack, as well as to develop a structured SE attack ontological model. The next section utilises these taxonomies and discusses a classification of an SE attack based on the type of communication that is employed.

2.5 SOCIAL ENGINEERING ATTACK CLASSIFICATION

A SE attack, as depicted in Figure 2.1, can be divided into two main categories: An indirect attack and a direct attack.

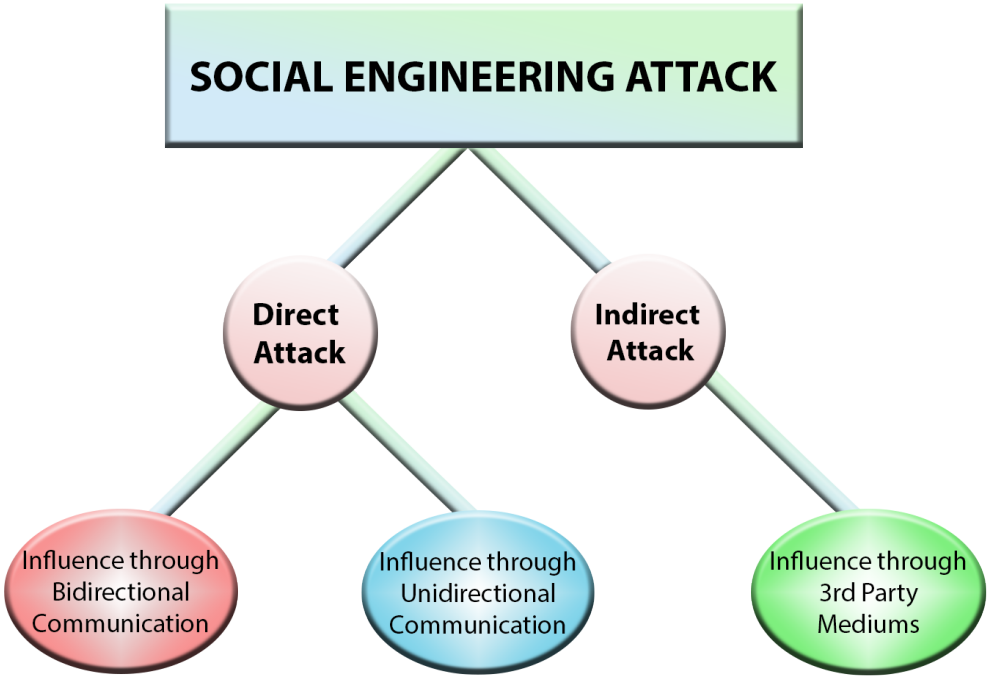

Figure 2.1: Social Engineering Attack Classification

An indirect attack refers to an incident where a third party medium is used as a way of communicating. Third party mediums typically include physical mediums such as flash drives, pamphlets or other mediums, such as web pages. Communication occurs through third party mediums when a medium is accessed by a target, without direct interaction with the social engineer.

A direct attack is an incident where two or more people are involved in a direct conversation. This conversation can either be one-sided or two-sided. Due to this, this type of attack is further classified into two ways of communicating: Bidirectional or unidirectional communication.

Bidirectional communication is when two or more parties take part in the conversation, in other words, a two-way conversation occurs. Each party consists of an individual, a group of individuals or an organisation. A popular example of an attack in this category is an impersonation attack, where the social engineer impersonates the target in order to gain access to something which the target has access to.

Unidirectional communication is a one-sided conversation where the social engineer communicates with the target, but the target has no means to communicate back with the social engineer. This is

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

The rest of this subsection explains the different categories, bidirectional communication, unidirectional communication and indirect communication, in more detail with an example of each. Each example discusses the various parts of an SE attack, as defined in section 2.3: a social engineer, a target, a medium, a goal, one or more compliance principles and one or more techniques. Compliance principles are principles used by the attacker, aided by different techniques, in order to persuade the target, through some medium, to comply with a request.

Bidirectional communication (Figure 2.2) is defined as a two-way conversation between two people. In the bidirectional communication category, the social engineer can either be an individual or a group of individuals. The target of the attack can be an individual or an organisation. The mediums that are frequently used for bidirectional communication are e-mail messages, face-to-face conversations or telephone conversations. Any compliance principle, technique and goal can be used in combination with a bidirectional communication medium.

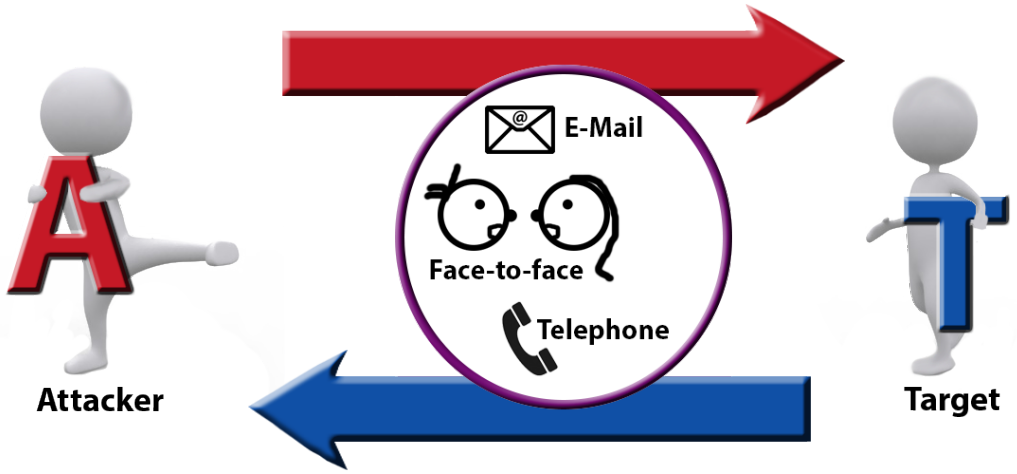

Figure 2.2: Bidirectional Communication

An example of an SE attack that uses *bidirectional communication* is one where a social engineer attempts to influence a call centre agent into divulging sensitive information regarding a specific client. In this example, both the attacker and the target are *individuals*. *Pretexting* is used as the technique for this attack because the social engineer impersonates the client whose information the social engineer

wishes to obtain. The compliance principle used in this example is *authority*, because the client impersonated by the social engineer acts as if he or she has authorised access to the information. The goal of the attack is to gain *unauthorised access* to the client's sensitive information.

Unidirectional communication (Figure 2.3) is very similar to bidirectional communication, except that the conversation only occurs in one direction: From the social engineer to the target. The social engineer and the target can either be an individual, a group of individuals or an organisation. The mediums that are frequently used for unidirectional communication are one-way text messages, e-mails or paper mail messages. Any compliance principle, technique and goal can be used in combination with unidirectional communication.

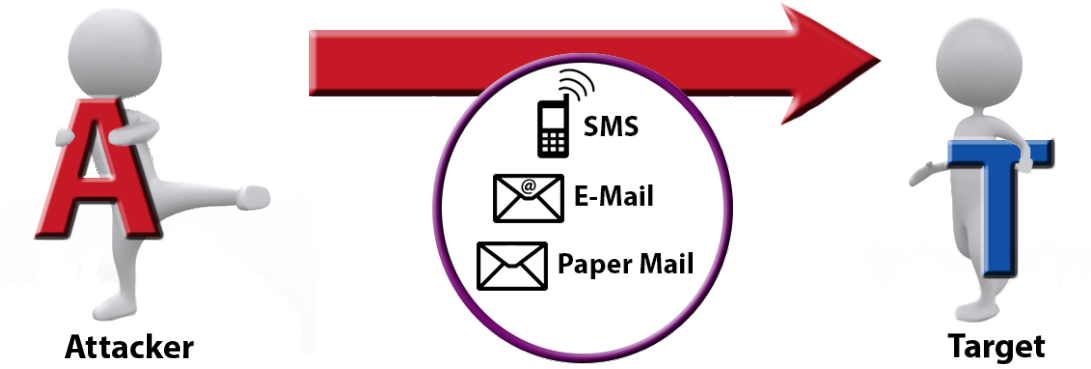

Figure 2.3: Unidirectional Communication

An example of an SE attack that uses *unidirectional communication* is an e-mail phishing attack where the target places an online order at some online store and waits for delivery of the item. The phishing e-mail is masked as an e-mail from the online store informing the target that a limited offer is available relating to the order. The target recognises the link between the e-mail and his order and clicks on the infected link. The target is specifically chosen. *Phishing* is the SE technique used for this attack and *scarcity* is the compliance principle. Since the e-mail states that it is a limited offer, the target feels that he or she has to explore this limited opportunity before it becomes unavailable. The infected link gives the social engineer *unauthorised access* to the target's computer.

Finally, there is **indirect communication** (Figure 2.4) which is defined as communication through a third party medium. The social engineer and the target can be either an individual, a group of individuals or an organisation. The mediums that are frequently used for indirect communication are

pamphlets, flash drives and web pages. Any compliance principle, technique and goal can be combined with indirect communication.

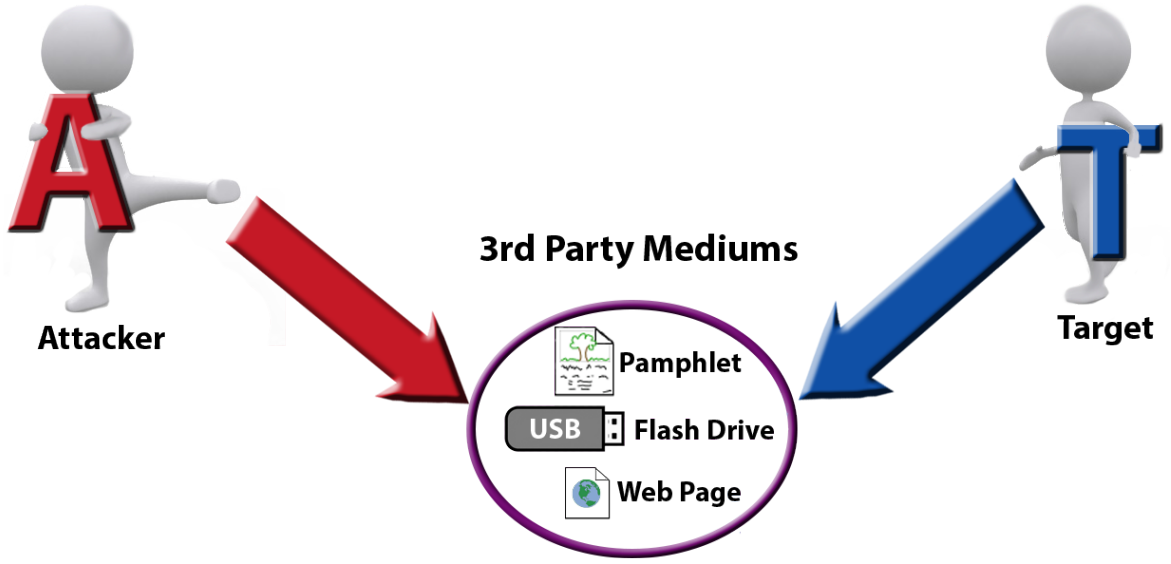

Figure 2.4: Indirect Communication via 3rd Party Medium

An example of an SE attack that uses *indirect communication* is when a social engineer leaves an infected flash drive lying around in a specifically chosen location with the intention of it being picked up by the target. The infection vector on the flash drive opens up a backdoor on the target's computer when inserted into the computer, allowing the social engineer unauthorised access to the computer. In this example the social engineer, as well as the target, are *individuals*. The technique used for this attack is known as *baiting* because a physical object is left in visible view of a target. The success of the attack relies on the curiosity level of the target. The compliance principle used is *social validation*, which states that someone is more willing to comply if they are performing some action they believe to conform to a social norm. The target may feel socially obliged to attempt to find the owner of the lost flash drive. This leads to the target plugging the flash drive into his or her computer which then activates the backdoor and unknowingly grants access to the social engineer. The goal of the attack is *unauthorised access* to the target's computer.

This has now provided one with the three different types of communication that can be utilised during social engineering attacks. These types of communication, alongside the social engineering taxonomies and definitions, are utilised to develop an ontological model of a social engineering attack in the following section.

2.6 SOCIAL ENGINEERING ATTACK ONTOLOGICAL MODEL

The model that is presented in this section has been compiled from various other taxonomies and the author's classification of SE attacks. The purpose of this ontological model is not just to define the domain but also to form the foundation for an ontology for SE attacks. The argument is that a taxonomy is too limited to define SE and SE attacks sufficiently. An ontological model provides additional structure to fully define this domain. According to Van Rees (2003), a taxonomy is a hierarchical structure to aid the process of classifying information, while an ontology is a well-defined set of definitions that create a taxonomy of classes and the relationships between them. Van Rees also states that "an ontology resembles both a kind of taxonomy-plus-definitions and a kind of knowledge representation language." [53].

It is clear from the other taxonomies discussed previously, that their authors tend to mix techniques, compliance principles, mediums and phases of an attack. The ontological model of the author represents each entity of an attack as well as the relationships between entities.

An ontology allows a formal, encoded description of a domain. All the relevant entities, their attributes and their inter-relationships can be defined and represented in a machine-readable model. Gruber (1993) defines an ontology as "formal, explicit specification of a shared conceptualisation." [54]. Noy and McGuinness define an ontology as: "... a common vocabulary for researchers who need to share information in a domain ... includes machine-interpretable definitions of basic concepts in the domain and relations among them." [55]. Ontologies have automated reasoning facilities that enable the derivation of new information from the facts contained in an ontology.

The model is based on the consolidated definition of an SE attack as depicted in Figure 2.5. The author defines a *Social Engineering attack* (Section 2.3) to have:

- one *Social Engineer*;
- one *Target*;
- one or more *Compliance Principle*s;

- one or more *Techniques*;
- one *Medium*; and
- one *Goal*.

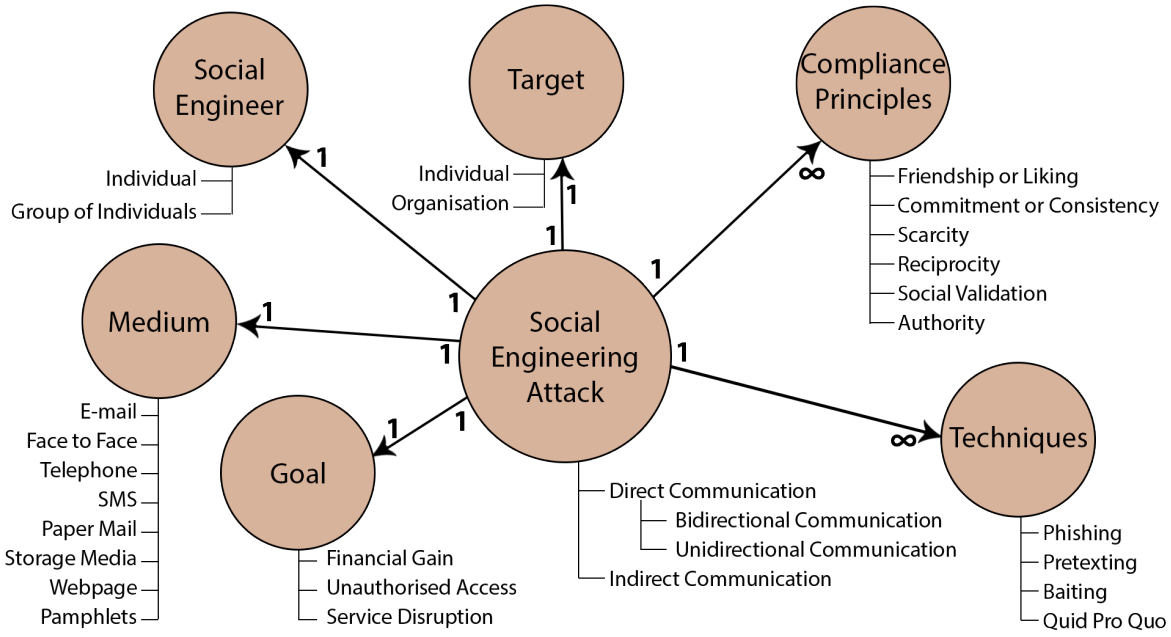

Figure 2.5: An Ontological Model of a Social Engineering attack

Each of the six entities is represented as a different class in the model. The subclasses of each class are shown in Figure 2.5. For example, the *Social Engineering Attack* class has two subclasses: *Direct Communication* and *Indirect Communication*. In turn *Direct Communication* has two subclasses: *Bidirectional Communication* and *Unidirectional Communication*.

The model, in its current state, provides the building blocks to further expand the model into a full ontology. As a future task, when the ontology is built from this model, additional relationships between these classes can be developed and described in detail. One example of a relationship between two classes is *performsAttack* between the *Social Engineer* class and the *Target* class. The model partially represents the proposed definition of *Social Engineer* (Section 2.3): *An individual or group who performs an act of Social Engineering*. The latter part of the definition requires representation of the verb *social engineer* and will be presented in the ontology as the relation *performsAttack*. The components of the model is further described in Chapter 5.

The following section concludes this chapter by providing an overview of the definition of social engineering, the social engineering attack classification and the ontological model for social engineering attack.

2.7 CONCLUSION

Organisations usually employ advanced technical security measures to minimise opportunities for unauthorised individuals, however, every organisation has employees who are likely to be susceptible to SE attacks. As electronic computing devices become more prevalent, the group of individuals who can be targeted by social engineering is increasing significantly. These reasons motivate why SE is such an important field of research. Although SE is a discipline that enjoys increasing attention, it is still not well defined. This chapter provides an overview of several definitions from the literature and shows that many researchers define SE to suit their specific topic of research.

In order for the field of Social Engineering to mature, it is required to have commonly accepted domain definitions. Based on all of the definitions and existing taxonomies that have been examined, this chapter proposes both a Social Engineering Attack Classification as well as a Social Engineering Attack Ontological Model.

The Social Engineering Attack Classification divides an SE attack into two classes: a direct attack and an indirect attack. The direct attack is further subdivided into an attack utilising bidirectional communication and an attack utilising unidirectional communication. The indirect attack class is further defined as an attack utilising third party mediums as a communication platform.

The Social Engineering Attack Ontological Model expands on the Social Engineering Attack Classification by providing six entities of an attack as well as the relationships between these entities. This model currently represents the definition of an SE attack and partially represents the definition of a social engineer.

In summary, this chapter is able to provide definitions for several terms within the domain of Social Engineering. The most important of these terms are Social Engineering and Social Engineering attack. The first one is defined as: *The science of using social interaction as a means to persuade an individual*

or an organisation to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity. The latter term is defined as: *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*.

For the purposes of this thesis, the objective addressed in this chapter was to introduce the reader to the concept of social engineering and social engineering attacks. This chapter provides the building blocks for the rest of the thesis. Several of the concepts introduced in this chapter is revisited throughout the thesis and expanded upon.

One item that was omitted from this chapter was the detection of social engineering attacks. Due to the fact that this thesis focuses on detecting social engineering attacks and the development of a SEADM, the entire following chapter is dedicated to discussing the detection of social engineering attacks.

CHAPTER 3 INITIAL SOCIAL ENGINEERING ATTACK DETECTION MODEL

3.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the required literature to understand the current state of social engineering attack detection. This chapter is based on a conference paper entitled "Social engineering attack detection model: SEADM" which has previously been published by the author at the Information Security for South Africa conference in 2010 [44]. This conference paper was the initial idea behind this thesis and this work is included to provide the reader with an overview of the first iteration of the SEADM. The author is aware that this section is not limited to a literature review, as most of the work here is the initial contribution towards the SEADM. A conscious decision is made to include this work in the literature section as it provides the reader with insight as to how this work evolved over the years. This chapter is merely the starting point of a larger research topic. This chapter is included to provide the reader with a look at the state of the social engineering domain in 2010 when the work was started. It also provides the reader with a frame of reference as to what social engineering attack detection entails. It is important to have this chapter included here, since the following chapter discusses ethical concerns regarding social engineering research; work that was also performed due to the manipulative nature of social engineering attacks and detection of these attacks.

3.2 INTRODUCTION

As clearly stated by various authors [56, 5, 57, 58], the human element is the 'glitch' or vulnerable element within security systems. It is the basic 'good' human nature characteristics that make people vulnerable to the techniques used by social engineers, as it activates various psychological vulner-

abilities, which could be used to manipulate the individual to disclose the requested information [56, 58].

Individuals make themselves even more vulnerable to social engineering attacks by not expecting to ever be a victim of such an attack, and many will never know that they were a victim of such an attack [5]. The majority of the public are not aware of social engineering, and do not fully comprehend the extent to which these techniques can cause them to divulge sensitive information. The divulged sensitive information may cause dire personal, economic and social consequences and losses for the individual. An individual may believe that the information they posses is of no particular value to another person, nor could it be used for any malicious act, and will thus be more willing to disclose information freely. However, the social engineer is dedicated to researching various aspects and gathering information from various sources. Combined the acquired information can have dire consequences.

On the other end of the spectrum, the individual might believe that they will not fall prey to such an attack, as they would be able to recognise such an attack instantaneously [5]. However, the social engineer is a skilled human manipulator, preying on human vulnerabilities using various psychological triggers that could foil human judgement.

The problem lies in successfully detecting social engineering attacks whilst working in a stressful environment where decisions must be made instantaneously [4]. It is for this reason that a practical model, that is easily implemented and used by all levels of employees, is necessary. This chapter provides the very first SEADM that was proposed, which formed the basis to why this thesis was written. This model was proposed to be used in combination with training on various social engineering techniques, the psychological vulnerabilities it may elicit, and institutional policies and procedures.

The two main perspectives of social engineering - the psychological perspective and the computer science perspective - are accounted for within this model. The psychological perspective focuses on the emotional state and cognitive abilities of the individual. The computer science perspective focuses more on information security areas and why the information should be deemed as sensitive. Other important factors incorporated within this model are the urgency of requested information and an individual's comprehension of the requested information.

The remainder of the chapter is constructed as follows. Section 3.3 provides background about social

engineering attacks from a purely psychological perspective, and Section 3.4 discusses background on the process of human reasoning and decision-making. Section 3.5 introduces the first iteration of the model, which was developed for social engineering attack detection and provides an in-depth discussion of each of the pertinent elements of the model. Section 3.6 provides scenarios to demonstrate the effectiveness of the model. Finally, Section 3.7 concludes with a summary of the SEADM.

3.3 PSYCHOLOGY BEHIND SOCIAL ENGINEERING

Various psychological vulnerabilities and triggers, used by social engineers, have been identified, which aim to influence the individual's emotional state and cognitive abilities in order to obtain information. To successfully defend against these psychological triggers, the individual will need to have a clear understanding of these triggers in order to recognise each during a social engineering attack. Seven psychological vulnerabilities has been defined by [59]. These psychological vulnerabilities are [5, 11, 56, 60]:

Strong Affect: When a strong emotion is triggered, such as anger, excitement, fear or anxiety, an individual's cognitive ability may be seriously hampered. This may include their ability to make decisions rationally, evaluate the situation, make counterarguments, and reason logically, making this a very effective technique used by social engineers [60]. A phishing attack could be used as an example. These are thoroughly planned criminal attacks, where websites are designed to masquerade as the authentic site, in order to obtain another individual's authentication credentials and confidential information illegally for financial gain. E-mail communication is one of the easier methods of reaching a large distribution of the population resulting in phishing attacks being mostly executed via this route in order to ensure the success of the attack. A disparity is created between the individual's perception and the truth, eliciting a heightened fear response, where cognitive abilities are compromised, and the probability of ensuring that the correspondence is legitimate will be minimal [61].

Overloading: This technique has a time element, with the result that the individual becomes cognitively pacified or compliant, through the bombardment of a series of hurried persuasive axioms [60].

Reciprocation: "One good deed deserves another"; Social exchange theory states that individuals, on receiving a kind gesture from another, feels obligated to reciprocate with kindness. The social engineer might create a problem for the individual, only to fix it again, in order to make the individual feel obligated to reciprocate by disclosing information [11].

Deceptive Relationship: To obtain information, the social engineer will identify an individual to purposefully build and establish a relationship. This is done with a particular purpose, as individuals tend to share information freely within established relationships [60].

Diffusion of responsibility and moral duty: The individual is made to believe that their actions - to disclose information, even though it is against policy - will have greater benefits and important beneficial consequences, such as to help save an employee or helping the institution, and that they will not be held solely responsible for their actions [60].

Authority: By the social engineer portraying an authority figure, the individual is more likely to comply with the request to disclose information, as an authority figure almost implicitly elicit a conditioned response to adhere to their wishes and demands, combined with a fear of punishment if the individual may appear to undermine their authority by verifying their legitimacy [11, 60].

Integrity and Consistency: Individuals have an intrinsic desire to uphold their commitments, even if it were not their own [60].

These triggers can be used to perform a social engineering attack on an unsuspecting victim, which can lead the victim to experience a sense of discomfort, whether just an uneasiness or even anxiety, as all these attacks prey on the victim's psychological vulnerabilities. One would expect that a victim would be able to use these clues of discomfort to detect that he is being targeted by a social engineering attack. However, this is the ideal and not reality. The human reasoning and decision-making process is extremely complex, and prone to error, hence the inability to detect these social engineering attacks themselves. In order to shed more light on this error-prone behaviour of humans, the following section discusses the human reasoning and decision-making process and how it applies to detecting social engineering attacks.

3.4 HUMAN REASONING

The human ability to make conscious, rational judgements, which underlie their decisions, will not always be the ideal. This can be ascribed to various human factors, such as limited information-processing capacity, the use of heuristics (mental processes, or shortcuts, used to simplify the process of judgement, which can lead to judgemental error), personal preferences, and a vulnerability to be influenced by emotions and manipulated by others. Human decision making is a complex process, where most decisions that need to be made will not have only one ideal option, and the same decision will not be made by all people [62, 63].

Within the subjective utility theory, the subjective experience of an individual is taken into consideration, where the goal is to maximise gain and to avoid losses. This subjective experience refers to the individual's own personal judgement on value (utility) and likelihood (probability), instead of objective criteria and computations, where personal characteristics have an impact [63, 64].

The individual will follow a series of steps to come to a decision. First, for each option, they will multiply the subjective probability by the positive subjective utility, followed by subtracting the calculation, as before, for negative subjective utility. Based on these expected values, individuals will make their decision [63].

Risk will always be an integral part of decision-making, as the possible outcome is uncertain. The subjective expected utility theory is the most widely applied model regarding risk decisions. Within this extended version of the subjective expected utility theory, it allows for subjective probabilities, where judgements are made based on the person's belief on likelihood, and where no objective mathematical probabilities are available. This theory cannot, however, predict human decisions. As indicated by the term subjective, each person will have their own set of values and characteristics. By considering the particular individual's subjective expected utilities and their subjective estimates of probabilities of cost and benefits, one can predict the optimal decision for that particular individual [62, 63].

Within this subjective expected utility theory model it is believed that the individual will try to achieve a well-reasoned decision by considering all the possible alternatives and information available, calculating the probability of each probable outcome, and the cost and benefits it may hold [62]. Based on this theory, a decision to disclose information was based on risk-benefit analysis [62].

Decision analysis, a technology based on subjective expected utility theory, attempts to aid better decision making [62]. This approach attempts to aid people to comprehend and have clarity regarding their goals and values, to search for possible options and verification of facts. One of the techniques used by decision analysis is decision trees. Decision trees are representations of decisions, which aid complex decision-making by breaking it down into more manageable components. Values are assigned to each element, whereupon ideal decision principles are applied to integrate these elements. By combing the probabilities and the utilities that correspond to each possible outcome, the best alternative is selected [62].

People do not possess a stable set of pre-existing values that are simply applied; their decisions will be determined by the present context, and the demands of the decision [62].

As indicated by literature, individuals find it difficult to make rational decisions in a limited time frame, especially regarding complex matters. With the skill of the social engineer and the complexity of the attack he is performing, at best, an uninformed individual would only be able to make an educated guess regarding the likelihood of being targeted by a social engineering attack. An individual would need a predefined set of guidelines on which to measure the likelihood of a social engineering attack in order to make a more informed decision.

It has now been shown that the human reasoning process is inherently flawed and that some form of intervention is required to aid the human reasoning process. The following section is devoted to proposing a practical application model to determine if a social engineering attack is being performed.

3.5 SOCIAL ENGINEERING ATTACK DETECTION MODEL

As indicated, a model is needed as guideline to detect social engineering attacks. The author proposes the SEADM, making use of a decision tree, by breaking the process down into more manageable components, and guidelines to aid decision-making in Figure 3.1.

This chapter firstly addresses each of these states individually as shown in Figure 3.1 before the full model is discussed with examples. Throughout this discussion the term individual is defined as the

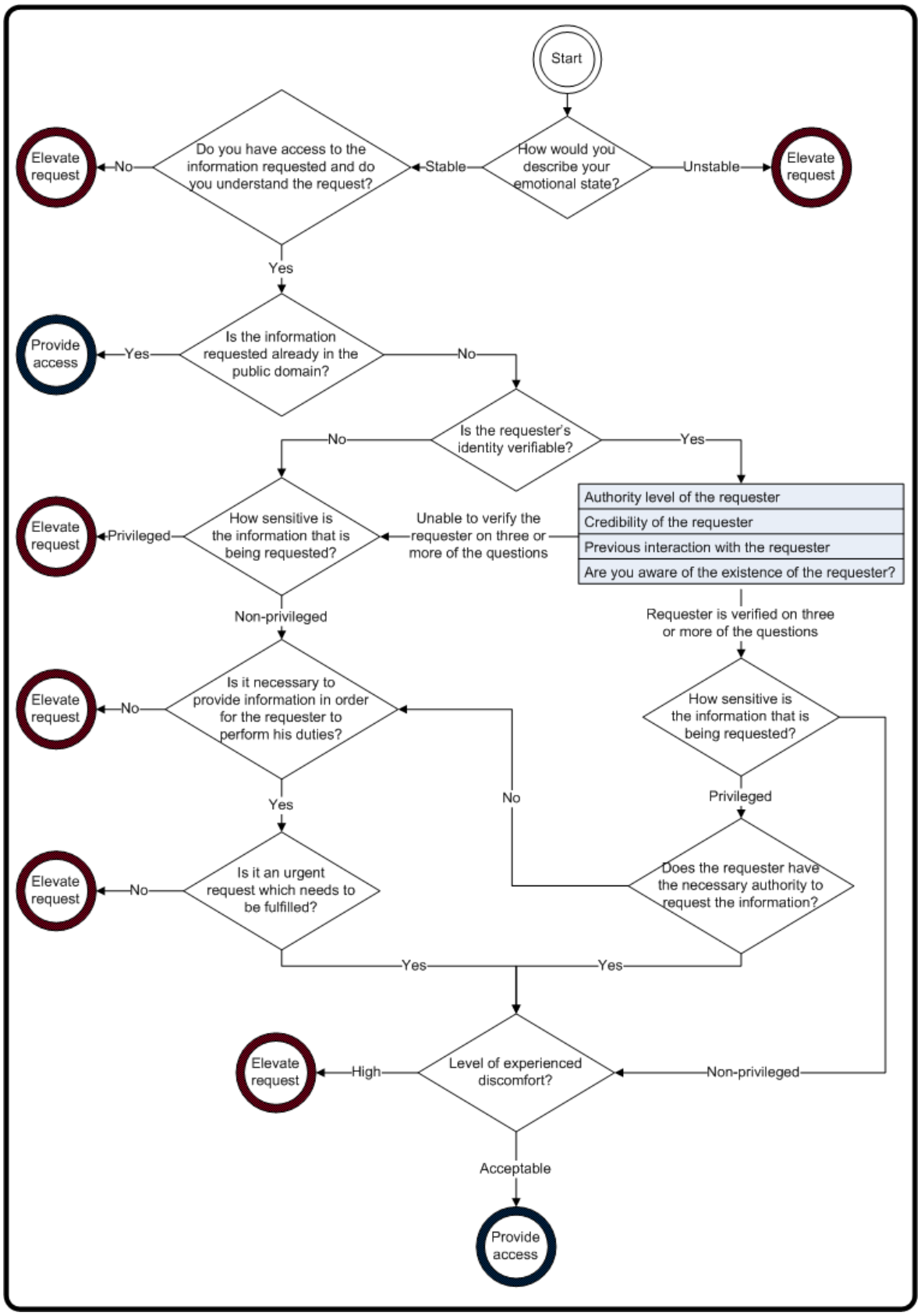

Figure 3.1: Social Engineering Attack Detection Model (Preliminary Version)

person dealing with the incoming call and the term requester is defined as the person who is making the call and requesting the information.

3.5.1 How would you describe your emotional state?

The first necessary step in this model, and one that will have to be considered throughout the process, would be for the individual to be conscious of, and evaluate their emotional state on an ongoing basis. This implies a consciousness of emotion and how it can affect one's decisions.

In the same manner, the individual should evaluate the emotion the requester elicit within themselves, as the psychological vulnerabilities, that might be triggered by a social engineering attack, is directly aimed to create certain emotional states in order to obtain information.

We are all familiar with a day that start off horribly and seem to continue with every possible thing going wrong. For example, the car broke down on the way to work, followed by a negative emotional experience whether it be family problems or an argument with a spouse or colleague. All factors and negative events influence our emotional state and hamper our ability to make rational, thought-through decisions [65]. In such a negative emotional state it is more likely to be a victim of social engineering: concentration is low, irritability and frustration is high, and an individual can possibly provide a requester with information just to get rid of them.

It is necessary to emphasise again what a critical role an individual's emotional state can play in the safekeeping of privileged information. If an individual is in a negative emotional space, the individual will not always be able to make a rational decision on the sensitivity level of the information of a request, or to whom it may be disclosed. This can result in costly losses to the institution and individual.

Awareness and consciousness of one's emotional state will not be an easy or even a possible task for all individuals. With training and rehearsal, this skill can and will improve. For this reason, the author is in the process of developing a quick self-evaluation electronic questionnaire that individuals will be able to use. However, in combination with the model, training by the institution can be emphasised on

the various techniques used, the psychological vulnerabilities the attacker may elicit, and institutional policy and procedures.

It is important to note that judging one's own emotional state could be a tedious matter and some individuals are unable to perform this task. It is for this reason that an automated self-evaluation electronic questionnaire would be implemented. The questionnaire will consist of a large database of questions, of which only a few would be used per evaluation of this state, as there is a time constraint associated with the model, making individuals unable and reluctant to perform a self-evaluation task taking an excessive amount of time. The time-frame for completing this state should be within a few seconds.

If the individual or the self-evaluating questionnaire finds that the individual is too emotional, the call or email request should rather be re-directed to another individual. Of course this has the implication, and danger, of people using this as a tool to shift their work responsibilities to another, promoting further frustration for all individuals involved. However, the dangers of social engineering, obtaining privileged information that can lead to great losses to the institution and possibly the individual, are a much greater threat.

3.5.2 Do you have access to the information requested and do you understand the request?

When a request for particular information is made, the individual needs to judge if they possess adequate knowledge regarding the requested information, and if they have access to the information which is being requested, to adequately provide this information. Obviously, if the individual does not have the knowledge required, they will not be able to provide the information, and could refer the requester to another individual, who will then also follow this model. If the individual judge that they have adequate knowledge on the subject in question the following step can be taken.

3.5.3 Is the information requested already in the public domain?

Individuals should have a clear understanding of what information are readily accessible to the public regarding their institution and related information. The information in the public domain can include

contact details and working hours, which can be available on the institutions website, and thus be legally provided to a requester.

3.5.4 Is the requester's identity verifiable?

The individual now needs to verify the identity of the requester, to enable them to make an informed and rational decision if information should be provided to the requester at a later stage of the model. If the requester's identity cannot be verified, a different set of states will be examined to determine if the information should be provided.

Important to remember is that the social engineer might be portraying himself as an authority figure within the institution, a computer technician, or any other persona that might elicit compliance. As humans we are inclined to make quick assumptions regarding people and their stature, based on trivialities such as clothing. If someone is dressed in the proper attire, use the appropriate institutional jargon, using an important individual's name, does not necessarily indicate that the individual is trustworthy. Social engineers do an enormous amount of research before an attack, if warranted. If at any time the individual feels unsure, they should contact their manager, to obtain authority to provide, or not provide, the information requested.

To verify the requester's identity, the following should be taken into account and used to form a global impression to base the decision on whether to provide or not provide the requested information: authority, credibility, previous interaction, and knowledge of the person's existence will have to be taken into account.

Some of the techniques that can aid in the verification process of an individual's identity are the following: Caller Identification; Calling back the requestor on a predetermined phone number; To request a secure e-mail; To request a secure password; To request a face-to-face interaction with the individual where he would provide proper identification, Where another employee can vouch for the requester; To contact the requester's immediate supervisor in order to verify his/her identity; To use an employee directory [5].

In this model it is suggested that the individual should be able to determine at least three of the four

components to successfully verify the individual. Each of these qualities will now be individually addressed.

3.5.4.1 Authority level of the requester

Authority is part of any institution, with an almost conditioned response from employees to adhere to their wishes and demands, combined with a fear of punishment if the individual may appear to undermine their authority [11]. For this reason it is a very effective technique used by social engineers to obtain privileged information. The institution needs to provide an environment where the employee feels comfortable, and are expected to question the authority figure's identity when disclosing sensitive information.

With determining authority, the employee also needs to know, with the help of a clear institutional policy, what authorisation level a particular person of authority has, with regards to what privileged information can be provided.

3.5.4.2 Credibility of the requester

The employee needs to judge the level of credibility of the requester. However, this is a challenging task, as establishing credibility is the first step the social engineer undertakes, and what the attack will be based on.

If the requester knows the jargon used by the particular institution, people easily assume that the requester is an employee at their particular institution. The requester could, for example, be an ex-employee, quite knowledgeable about the jargon and procedures. Such ex-employee might seek revenge with the goal of obtaining particular sensitive information. The credibility of the requester is measured on the basis of how he/she responds on predefined of set of questions which can be used to determine the credibility of a requester.

3.5.4.3 Previous interaction with the requester

If the individual had previous interaction with the requester, especially a longstanding history of interaction, the decision and knowledge to what information can be provided will be an easy task. However, few interactions with the requester, especially by telephone and email alone, should be considered in conjunction with other verification techniques, to be able to make an informed and safe decision regarding the disclosure of information.

3.5.4.4 Are you aware of the existence of the requester?

This refers to the knowledge that the requester exists within the institution or an outside collaborating partner on a project can support the verification of the requester. However, this should also be used in conjunction with the other verification techniques, as the requester could be a social engineer portraying himself as the well-known figure in order to obtain privileged information.

It is suggested that within institutional policies and procedures, a classification system of information should be established, whereupon a document of all personnel should be compiled and made freely available to employees, indicating what level of information authorisation each has in order to simplify the process.

3.5.5 How sensitive is the information that is being requested?

It is critical that the individual are knowledgeable, and have absolute clarity, what information is privileged, and what information are authorised to be provided, and to whom, thus depicting the level of information sensitivity. This skill can be nurtured and enhanced through training on institutional policies and procedures.

For the purpose of this model, information is divided into two categories, privileged and non-privileged information. Privileged information indicating information requiring a form of authorisation, and non-privileged information indicating information that requires no authorisation and are freely available. The proposed model should be used in conjunction with an institution's policies and procedures on information sensitivity. These policies and procedures should include clear, easily understandable and

easily accessible guidelines to verify the authorisation level needed in order to request the specific information.

As each institution is unique, each will have to create and establish their own security policies to address the classification of the sensitivity of particular information, under which circumstances it may be divulged, and to which individuals or institutions. These policies should also include processes and accountability for reporting suspected incidents [11].

After determining whether information is privileged or non-privileged, the individual will need to determine if the requester has the necessary authority to request the information.

3.5.6 Does the requester have the necessary authority to request the information?

With the aid of the previous steps, the individual possesses the necessary knowledge regarding the requester's identity and authority level, together with the information classification. The individual can now determine whether the requester has a level of authority on the same level or higher as the level of sensitivity of the information. If the requester possesses authorisation on the same authority level or higher needed for the particular information, the next step - the level of experienced discomfort - can be considered.

However, if the authority figure does not have the necessary authorisation, or if the individual feels that the request made is not legitimate, the model will treat the requester as a non-verified individual. In this scenario the following step - to determine the necessity of the information to fulfil required duties - will be considered.

3.5.7 Is it necessary to provide information in order for the requester to perform his duties?

A subjective estimation needs to be made if it will be beneficial or detrimental to provide the information to the requester at the particular time of the request, as well as how it could empower the individual to complete their work. The individual should be sure that if he/she provides information to the requester that it would indeed be beneficial to both parties involved.

Apart from establishing if the information would help the requester to complete his duties, one would also need to consider the urgency of the request.

3.5.8 Is it an urgent request which needs to be fulfilled?

The individual needs to assess the urgency that the requested information is needed. If the information is not urgently needed, and any doubts exist, the information does not have to be provided, or can be provided at a later time. With the time leniency, authority can be consulted, who can choose to further investigate, or provide authorisation to divulge the requested information.

If the information is urgently needed, whether it is to complete an urgent project, or in a life threatening situation such as where an individual's medical insurance number is required due to an injury on duty, the employee should consider the next step of level of experienced discomfort.

3.5.9 Level of Experienced Discomfort

Evaluation of one's emotions is again emphasised, where an individual will have to trust the emotions they are experiencing at that particular time, e.g. "trust your gut". If the level of discomfort experienced is evaluated as too high, information should rather not be provided, as certain techniques used by social engineers may elicit high levels of emotional discomfort, enabling them to obtain privileged information. Part of the social engineer's skill set is the ability to profile individuals, using the appropriate technique for the particular individual, forcing them into a desired role. This technique is called alter-casting [4]. In a certain scenario they may be aggressive and threatening towards the individual, causing high levels of anxiety, where the individual's cognitive ability to reason, to be able to stay calm and focused, and to be able to make rational counterarguments, are detrimentally influenced. In another scenario, and also the most frequently used form of this technique, the individual will be ascribed to the role of helper, where the individual could experience discomfort and possibly guilt - an emotion most people try to avoid - if they do not oblige to the request.

If, however, the individual does not experience any discomfort or if the level of discomfort is understandable and acceptable, information can be provided, as the previous steps have been successfully completed.

This concludes the the discussion on the SEADM and what each state entails. The next section demonstrates the application of the model by use of examples.

3.6 USAGE OF THE SOCIAL ENGINEERING ATTACK DETECTION MODEL

Three example scenarios are provided within this section. The first scenario is a legitimate request by a bank account holder, requesting his bank account balance. The second scenario also depicts a request for a bank account balance, however, by a social engineer. The third depicts a basic scenario where a request is made regarding the closing time of a store. Within all the provided scenarios in this chapter, it will be assumed that the individual dealing with the request is in a stable emotional state.

3.6.1 Scenario One

A telephonic request is made to obtain a personal bank account balance. The process, according to the SEADM model, will be following:

- Emotional state of the call centre agent will be analysed, which will equate to stable.
- Do you have access to the information requested and do you understand the request? Yes.
- Is the information requested already in the public domain? An individual's bank balance is not public information and will, thus, be necessary for the agent to verify the identity of the requestor.
- The requestor will need to identify himself, and establish his credibility by providing the call centre agent with his personal information. The call centre agent then verifies the information by comparing it to the information on the system when the bank account was created. This verifies the question of being aware of the existence of this requester, as well as the authority level of the requestor.
- How sensitive is the information being requested? A bank account balance is classified as privileged information.

- Does the requester have the necessary authority to request the information? Yes.

- Lastly the call centre agent would need to analyse his level of experienced discomfort, which would be acceptable as there were no issues in this call.

Within the process completed, in this scenario, access can be provided, allowing the call centre agent to provide the requester with his bank balance.

3.6.2 Scenario Two

This scenario also depicts a request for a bank account balance, however, by a social engineer.

- Emotional state will be analysed, which will equate to stable.

- Do you have access to the information requested and do you understand the request? Yes.

- Is the information requested already in the public domain? An individual's bank balance is not public information, thus it will be necessary for the agent to verify the identity of the requestor.

- The requestor, who, in this scenario is a social engineer, will attempt to identify himself. This can proceed in one of two ways. The social engineer could be in possession of adequate information pertaining to the victim's personal and banking details. This information used in conjunction with his various skills and techniques, for example overloading, can convince the call centre agent he is the legitimate requester. This can lead the call centre agent to experience a high level of discomfort. The call centre agent can elevate the request to another individual with higher authority to adequately manage the request, or can deny access to the information. *To fully explain the model, this thesis will examine the alternative route, where the social engineer failed to validate himself as the owner of the bank account but he has validated himself as a friend of the owner of the bank account.*

- How sensitive is the information being requested? A bank account balance is classified as privileged information.

- Does the requester have the necessary authority to request the information? Within this alternative scenario, the answer will be no. A friend will not have authorisation to privileged information as a bank account balance.

- Is it necessary to provide information in order for the requester to perform his duties? The social engineer can portray himself as the bank account holder's accountant, explaining that he needs the information to complete his duties. Assuming the call centre agent allows this, he will move onto the urgency test.

- The call centre agent needs to determine the urgency of the request. However, a legitimate accountant would ask the account holder to contact the bank and obtain the necessary information. In this scenario the call centre agent would need to elevate the request, and report the request as a suspicious.

This scenario depicts how a social engineering attack could have been thwarted. The last scenario depicts a request to public information.

3.6.3 Scenario Three

Within this scenario a request is made regarding the closing time of the institution.

- Emotional state will be analysed, which will equate to stable.

- Does the individual have access to the information requested and understand the request? Within this scenario the individual have the necessary information regarding the operating hours of the institution and understands what information is being requested.

The operating hours of the institution is information which is already in the public domain, and thus can be provided to the requester. This chapter concludes by providing a brief summary and the potential advantages it may hold to an institution if applied together with adequate training.

3.7 CONCLUSION

Social engineering is very difficult to detect, as the social engineer possess various skills and effective techniques, preying on human vulnerabilities, which makes these attacks often go without notice. What makes detection even more difficult is that many people are unaware of this technique and the potential threat, and dire consequences it holds for the individual and for organisations.

As of yet, only training has predominantly been considered as preventative measure to social engineering. However, it has been shown that training is soon forgotten, especially in the real work environment, rendering training ineffective against social engineering. It is proposed that a visible, practically applied, user-friendly aid, such as the SEADM, will aid in the awareness of threats, protecting against social engineering.

It has been shown by the use of scenarios that the proposed model is indeed feasible as a preventative measure to social engineering attacks. This model already made a valuable contribution to the field of social engineering, as it aids in the detection of social engineering attacks, by breaking down the decision-making process into manageable components.

This model was the first of its kind to aid in the detection of social engineering attacks. The SEADM, in the state that it is proposed in this chapter, had many shortcomings though. These shortcomings were mostly due to the initial design of the model, when it was intended to only be used in a call centre environment. The first iteration of the SEADM only focused on bidirectional communication and did not cater for either unidirectional communication or indirect communication.

The goal of this chapter was to inform the reader of the need for social engineering attack detection based on the impact of human reasoning abilities. Furthermore, the initial SEADM is shown to the reader as it was the idea from which this thesis originated and formed the bases for further development. The following chapter focuses on the last part of the literature that was identified to be of vital importance, namely the ethical concerns regarding social engineering research. When research into social engineering was started for this study, there was no guideline with regards to ethics when it comes to social engineering research.

CHAPTER 4 ETHICAL CONCERNS REGARDING SOCIAL ENGINEERING

4.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the ethical concerns that needs to be taken into account whilst performing research in the social engineering field. This chapter is based on both a conference paper titled "Social engineering from a normative ethics perspective" and a journal paper titled "Necessity for ethics in social engineering research" which has previously been published by the author [46, 66]. The conference paper has been peer reviewed and was published at the Information Security for South Africa conference in 2013 [46]. The journal paper was a significant extension of the conference paper and was published in the Computers & Security Journal in 2015 [66].

Social engineering is deeply entrenched in the fields of both computer science and social psychology. Knowledge is required in both these disciplines on performing social engineering and hence the need for social engineering based research. Several ethical concerns and requirements need to be taken into account when social engineering research is conducted to ensure that harm does not befall those who participate in such research. These concerns and requirements have not yet been formalised and most researchers are unaware of the ethical concerns involved in social engineering research. This chapter is used to provide the reader with an overview of the ethical concerns regarding social engineering in public communication, penetration testing and social engineering research. Furthermore, this chapter also discusses the identified concerns with regard to three different normative ethics approaches (virtue ethics, utilitarianism and deontology) and provides their corresponding ethical perspectives, as well as practical examples of where these formalised ethical concerns for social engineering research can be

beneficial. This chapter is included in the literature section as it allows the reader to have an overview of the ethical constraints under which this research had to be performed.

4.2 INTRODUCTION

Social engineering attacks may have unintended after-effects on the victim. These may be so severe that they may, for example, lead to suicide [67] or other forms of trauma. The ethical concerns related to social engineering attacks, as well as the consequences of such attacks, could well be minimised if the right actions are taken after the attack.

When research in relation to social engineering is conducted on participants, several ethical requirements need to be taken into consideration. The problem is that these requirements have not yet been formalised and most researchers are unaware of the ethical concerns that affect social engineering research. This chapter aims to discuss the ethical concerns that need to be taken into consideration when social engineering is performed in a non-malicious fashion.

Non-malicious attacks are categorised according to the three different environments defined for this chapter in which attacks may happen, namely public communications (such as radio and television), penetration testing and social engineering research. Social engineering attacks performed in any of these environments are not intended to cause harm to the victim or to make malicious use of the information gathered in the attack.

The current research is important in the computer science domain as social engineering has very strong cross-disciplinary relations with social psychology [37, 44, 45]. Computer science researchers are not always aware of all the ethical concerns while dealing with human participants in a research study. Therefore research needs to be conducted on the ethics regarding social engineering to reduce and simplify the ethical constraints for a computer scientist involved in research.

The remainder of the chapter is constructed as follows. Section 4.3 provides a background about ethics and discusses ethics in terms of three main approaches to normative ethics. Section 4.4 introduces three chosen environments in which social engineering attacks can be performed. Section 4.5 lists and describes different social engineering ethical concerns framed in scenarios from each environment.

Section 4.6 discusses the ethical concerns presented in Section 4.5 in terms of three ethical perspectives. Section 4.7 provides the reader with practical examples of how this research can be beneficial and Section 4.8 contains a summary of the ethical concerns and concludes the chapter.

4.3 BACKGROUND ON ETHICAL PERSPECTIVES

This thesis focuses on three main approaches to normative ethics: virtue ethics, utilitarianism and deontology [68]. Normative ethics deals with the 'right' and the 'wrong' of interpreted social behaviour [69]. The main difference between these three perspectives lies in the way they approach a moral dilemma, and not necessarily in its consequences.

The next section discusses the three different approaches of normative ethics and how each ethical perspective is measured.

4.3.1 Virtue Ethics

Virtue ethics is defined as the ethics that emphasises the virtues, or moral character, of an individual's actions [70]. It focuses more on the character of the individual or the character's traits that guide the individual to his or her actions. 'Virtue', as defined in the Oxford Dictionary [71], is behaviour showing high moral standards and a quality considered morally good or desirable in a person.

In virtue ethics, morality is not measured by the rules and rights of the world. Morality is measured by the classic notion of the character, which includes honesty, fairness, compassion and generosity, to name a few. It focuses on the individual and not on the community [72].

Virtue ethics started in ancient Greece, but was revived as a rival account to deontology and consequentialism and their understanding of morality. Virtue ethics is self-centred and focuses on answering questions such as "How should I live?" and "What type of person should I be?" [73].

As a common test for virtue ethics one needs to ask the question: "Will doing this make me a better (or worse) person?" The truly wise person will know what is right, do what is good, and therefore be happy [74].

To apply virtue ethics to an ethical concern one needs to consider whether the act would be the kind of thing that a virtuous person would do [75]. A virtuous person is someone who displays the ideal character traits, such as always being kind to everyone in all situations because that is their character and not because it is required of them.

Since it is difficult to measure virtue ethics in a social engineering computer science domain, this thesis will use both the Institute of Electrical and Electronics Engineers (IEEE) code of ethics and the general moral imperatives from the Association for Computing Machinery (ACM) code of ethics. These ethical codes are most well known in the field of computer science research. This thesis uses examples that include individuals who do not subscribe to either the IEEE or the ACM code of ethics. The codes, however, exemplify what kinds of codes people who do value virtue ethics would subscribe to. Many of the professional requirements in these codes can be extended to any profession, and only some of them are specific to the Information and Communications Technology (ICT) field. Thus, for the purposes of this thesis, a virtuous person is seen as someone who complies with both the IEEE and ACM codes of ethics as provided in the excerpt below.

The IEEE code of ethics [76] states that: *We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:*

1. *to accept responsibility in making decisions consistent with the safety, health, and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;*

2. *to avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist;*

3. *to be honest and realistic in stating claims or estimates based on available data;*

4. *to reject bribery in all its forms;*

5. *to improve the understanding of technology; its appropriate application, and potential consequences;*

6. *to maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;*

7. *to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;*

8. *to treat fairly all persons and to not engage in acts of discrimination based on race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression;*

9. *to avoid injuring others, their property, reputation, or employment by false or malicious action;*

10. *to assist colleagues and co-workers in their professional development and to support them in following this code of ethics.*

The ACM code of ethics [77], under the general moral imperatives section, states that: *As an ACM member I will:*

1. *Contribute to society and human well-being.*

2. *Avoid harm to others.*

3. *Be honest and trustworthy.*

4. *Be fair and take action not to discriminate.*

5. *Honor property rights including copyrights and patent.*

6. *Give proper credit for intellectual property.*

7. *Respect the privacy of others.*

8. *Honor confidentiality.*

This thesis will use the following guideline in order to perform the test for virtue ethics: The social engineering attack that needs to be performed provides a window through which other people can see the social engineer for who he or she really is [78]. One then needs to examine the action from the perspective of these other people who are able to judge one's character from one's actions in order to measure virtue ethics. In the case where those other people would consider one to be a virtuous person in terms of both ethical codes, the attack will then be seen as ethical according to virtue ethics. The opposite is true if those other people would see one as a bad person for performing the social engineering attack.

4.3.2 Utilitarianism

Utilitarianism is the most common form of consequentialism. As in consequentialism, utilitarianism says that the rightness of an action is determined by the consequences of the specified action. Utilitarianist ethicists measure whether the action is ethical based on the outcomes of the action [68]. This approach involves analysing the impact of the individual's actions and the impact of this action on the majority of other people.

This can be either for the interest of the individual or for the majority of society [72]. For the purposes of this thesis, to test utilitarianism one needs to decide how it affects the majority of society. If the majority of society gains from the consequences, it is ethical, otherwise it is unethical.

To apply utilitarianist ethics to an ethical concern, one needs to consider the consequences of performing a social engineering attack on an individual and anyone else affected by the consequences of this attack. In utilitarianism, the consequences are assessed in terms of people's well-being. If the social engineering attack produces the best overall consequences for the community's well-being and the benefits to the community outweigh the consequence to the victim, then the utilitarian considers it ethically correct [79].

4.3.3 Deontology

Unlike the previous approaches, deontology focuses on adherence to the rules of the world in order to measure whether an action is right or wrong. It is also known as 'duty' or 'obligation' based ethics

[80]. Deontology focuses on the ethical act and some deontologists believe that there are universal rules regarding right and wrong behaviour [72]. Deontologists live in a world of moral rules, such as [80].

- It is wrong to kill innocent people.
- It is wrong to steal.
- It is wrong to tell lies.
- It is right to keep promises.

To test for deontological ethics, the basic rule "do onto others only that to which they have consented" [81] is followed. It is ethical if the individual is performing a morally right action, regardless of the consequences [80].

To apply deontological ethics to an ethical concern one needs to consider whether a social engineering attack would be conforming to moral rules that seem a priori logically correct. From a deontological perspective, the aforementioned rules need to be adhered to for the most part, irrespective of their consequences. If any part of the social engineering attack does not strictly adhere to the deontological rules, the entire attack can be seen as unethical. The opposite would be true when the social engineering attack adheres to all the deontological rules of the world.

The following section discusses three chosen environments in which social engineering attacks can be performed. It also shows how public communication and penetration testing fit in with social engineering research.

4.4 SOCIAL ENGINEERING ENVIRONMENTS

As mentioned earlier, this thesis focuses on three main environments in which social engineering can be performed, namely public communication, penetration testing and social engineering research.

These environments were selected as they provide the broadest base to identify specific scenarios in which social engineering attacks are performed.

Social engineering attacks performed with non-malicious intent are mostly performed in one of these three environments. The public communication environment involves public media where there is a presenter who may utilise manipulation techniques without realising that they amount to a social engineering attack. Penetration testing and research is more focused on using social engineering to help a third party discover vulnerabilities in their system and on the study of social engineering.

The following subsections describe each of the three environments in more detail.

4.4.1 Public Communication

In this environment, communication with the public occurs through some public communication medium such as radio or television. Social engineering attacks that happen in this environment are normally for the goal of entertaining listeners or viewers. The intent of these attacks is mostly not to harm the victim, although the harm can occur unbeknown to the presenter. The presenter may be unaware that he or she is performing a social engineering attack. The performed attacks can also have unintended harmful consequences.

4.4.2 Penetration Testing

In this environment, social engineering penetration tests are performed, which are designed to mimic attacks that actual malicious social engineers will use to steal data [82]. This can include attacks over the phone or the internet, but also on-site, by doing a 'break-in' into a physical place. The intent for these tests is not to cause harm, but rather to help improve the security by finding the vulnerabilities in the security system, whether physical or virtual.

The subjects who fall prey to the penetration test can feel guilty that they were not vigilant and this could lead to further personal harm. Management is required to view the penetration test report in an objective manner and not to take action against the employees who fall prey to the attack.

4.4.3 Social engineering research

Social engineering research constitutes a third environment in which social engineering techniques may be required. In this environment, social engineering attacks and SE awareness testing can potentially be performed as part of the research. Social engineering research involves a large environment and can also encapsulate the environments of penetration testing and public communication. This overlap of the social engineering research environment is depicted in Figure 4.1.

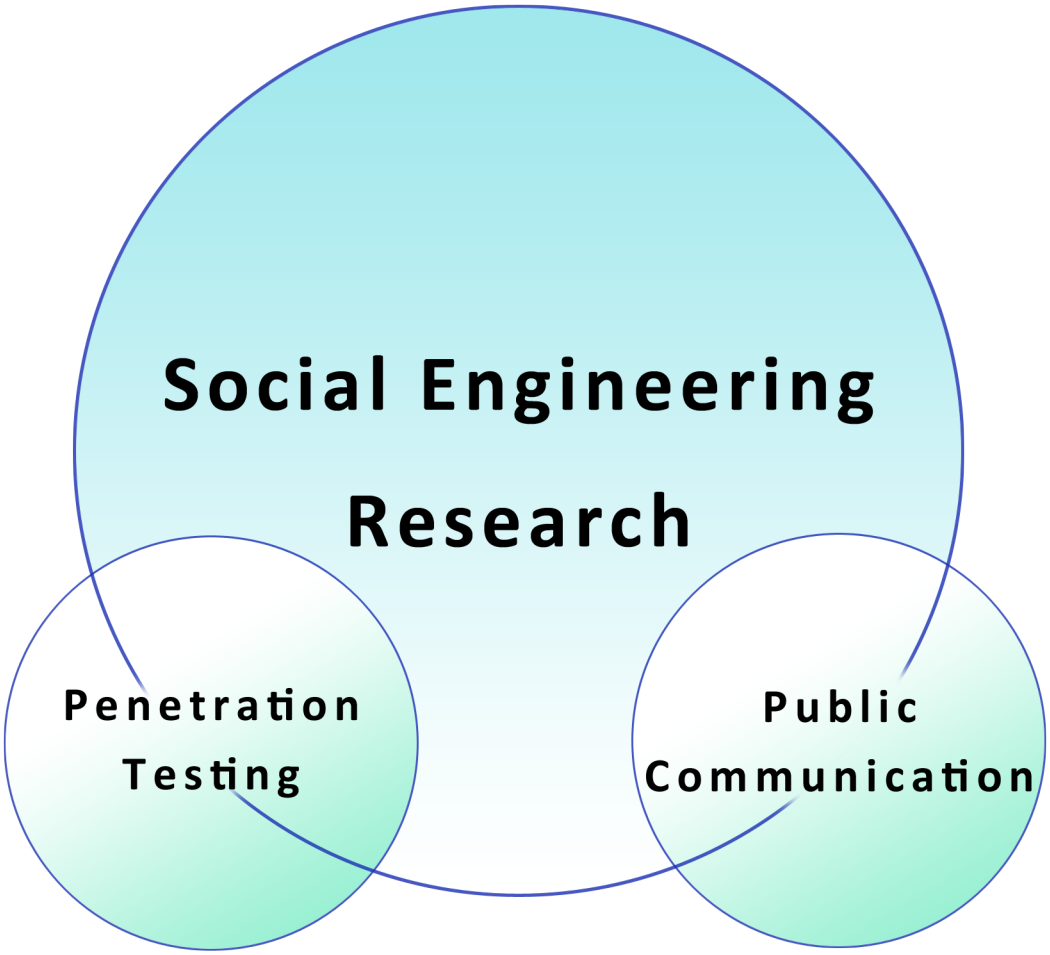

Figure 4.1: Overlap of the Social Engineering Research Environment

Social engineering research consists of several techniques that are required to gain accurate research results. In some of the research scenarios, the participants are required to be subjected to social engineering techniques without being requested to provide informed consent. The intent of this research

is not to harm participants, although they may deliberately be kept unaware of their participation in social engineering research. The reason why informed consent from the participant is not provided is because the participants will act differently if they are aware that they are participating in social engineering research and this may provide inaccurate results.

The next section lists some scenarios of social engineering attacks within these three environments. Ethical concerns related to each of the scenarios are also extracted.

4.5 SOCIAL ENGINEERING ETHICAL CONCERNS

Each of the aforementioned environments allows the researcher to provide several scenarios within context. The goal of the scenarios is to frame the different social engineering ethical concerns and to provide a context in which to examine them.

In each of these scenarios a single ethical concern is provided that relates directly to the specific scenario and environment. Not all of the scenarios adhere to the formal definition of social engineering and have been adapted to highlight the ethical concerns. The goal of this section is to provide the reader with all of the ethical concerns regarding social engineering, whilst providing a scenario in which to frame and later on discuss the ethical concern.

The subsections below list all of the ethical concerns and the specific scenario in which to frame them from the different environments.

4.5.1 Public Communications

The following scenarios are framed in the social engineering public communications environment.

4.5.1.1 Royal family scenario

The first scenario is based on an incident concerning the British Royal family [67] when Prince William's wife, Kate, was admitted to hospital. Two Australian DJs phoned the hospital pretending

to be Queen Elizabeth and Prince Charles, concerned about Kate's state of health. This prank was broadcasted live on the radio station to the public.

By using social engineering tactics and deceit, the DJs were able to talk their way through the hospital's switchboards and eventually got connected to Kate's personal nurse. Tactics included making noises in the background to simulate the queen's corgis barking. The DJs convinced the nurse to give them information regarding Kate's health status, which the nurse — believing it is the queen and the prince on the phone — gave out freely. The goal of this phone call was to provide entertainment and sensitive information to the radio station's listeners.

The ethical concern regarding this scenario is as follows: *Is it ethical to use social engineering to gain the trust of an individual?*

4.5.1.2 Radio prankster scenario

This scenario uses a radio prankster from a South African radio station, Highveld Stereo, Darren 'Whackhead' Simpson [83]. Darren is well known for the pranks he pulls on people for radio entertainment. Darren's career as a prankster has been so successful that he has published several audio prank collections discs that are sold all over the world.

In order to arrange a prank by Darren, someone who knows the victim sets up a prank with Darren after which he performs the actual prank. Darren thus has permission from either friends or family of the victim to perform the prank. Darren uses many social engineering techniques to convince the victim of his story, including background noises and different voices. He also gathers information about the person he is pranking beforehand in order to trick the person into believing his story. His intent is not to harm the victims, but to provide entertainment to Highveld Stereo listeners. After each prank, Darren reveals his identity and debriefs the victim.

The ethical concern for this scenario is as follows: *Is it ethical when delegated permission is used to perform social engineering techniques for public comical relief?*

4.5.1.3 Con artist scenario

The final scenario from the public communications scenarios is related to con artists on television. Several con artists hosting television programmes give audience members false hope by playing on their emotions, for example Marietta Theunissen from "Die ander kant" (The other side) [84]. She allows participants to 'speak' to their deceased relatives and states her field of work as "Psychic and Clairvoyant Readings, Health and Self-development".

Psychics such as Marietta, and Tarot card readers use psychological methods and information gathering techniques to play on emotions and make people believe they are speaking to their deceased loved ones, or to give them a glimpse into their future endeavours. They trick people into believing these things by using the victim's emotions as a method. The individuals want to believe that they have just spoken to their deceased family members and therefore the psychics can easily manipulate them into believing that they have indeed done that.

General phrases such as "Your grandmother is happy and loves you very much" are used, which can be applicable to almost any individual over a certain age. In the case where the victim has a deceased grandmother, the victim will have an overwhelming emotion and will believe that the phrase is aimed at him or her specifically.

In essence, these con artists give people a sense of false hope by tricking them with social engineering techniques. The intent is not to hurt these people, but to gain money or fame out of it. In reality, however, some people get hurt in the process as they can potentially suffer negative consequences from the advice provided by the con artist.

The ethical concern for this scenario is as follows: *Is it ethical to use information gathering techniques to provide participants with false information and to exploit them for either financial gain or fame?*

This concludes the ethical concerns related to the public communication environment. The following section deals with the penetration testing environment.

4.5.2 Penetration Testing

The following scenarios are all framed in the social engineering penetration testing environment.

4.5.2.1 Security guard scenario

This scenario describes a physical on-site penetration test and was adapted from a scenario discussed on a podcast presented by Chris Hadnagy [85]. At some physical location a security guard's duty is to patrol the premises. This involves the guard walking around the building constantly making sure no one gains illegal access to the building.

A penetration tester is hired to try and break into the building. The tester first observes the situation for a bit and finds the exact pattern of the guard patrolling the building. The tester then times the guard's patrol path and observes how much time will be needed to break in from the one side of the building while the guard is patrolling any of the other three sides. The penetration tester then uses lock picks to gain access to the door on one side of the building whilst the guard is correctly patrolling, as his job requires, the other three sides of the building.

The intent of the penetration test is not to cause harm to the security guard, but to provide management with a detailed report on how access to the premises can be gained. The intent of the penetration test is to show the organisation where there are vulnerabilities in their security. The penetration test is not specifically aimed at the single security guard but at the organisation as a whole.

The ethical concern regarding this scenario is as follows: *Is it ethical for the employee to bear the consequences of the successful infiltration, when the actual reason for the successful infiltration is not due to the employee's negligence?*

4.5.2.2 Generalised human vulnerability scenario

In this scenario a penetration tester is hired with the goal to gather sensitive information from an organisation. The penetration tester chooses an employee as his entry point.

The employee who has been chosen earns a minimum wage and is looked down upon by the general public due to his status. During the penetration test, the social engineer offers the employee a significantly higher sum of money than his or her salary in return for information over which the employee has control. Since the employee needs the money and is attracted to the big amount that is never reflected in his or her salary, the information is given to the penetration tester. The employee is also not necessarily aware of the sensitivity of the information that is given out.

Many people will fall for this type of an attack as the reward that is offered is unreachable by the victim under normal circumstances. If they are offered a sum of money that is significantly higher than their salary or any salary they might ever earn, most people will see it as an opportunity and ignore the danger in it. In the rare case where the victim sees the opportunity as dangerous, it should also take little convincing from a skilled social engineer to get the victim to cooperate.

The ethical concern regarding this scenario is as follows: *Is it ethical to exploit a personal weakness of an employee when it is known to be common human nature to fall prey to this type of attack?*

4.5.2.3 Receptionist scenario

In this scenario, a penetration tester is hired to attempt to gain sensitive information from the organisation. The penetration tester chooses the receptionist as a possible weak link in the organisation as it is the latter's duty to help customers, to the best of his or her ability, while the customer is in the reception area of the organisation.

The penetration tester therefore enters the reception area as a customer. While waiting for a scheduled appointment, he takes out his laptop and realises that he has no network access. He sees that there is a network access point that he can potentially use to connect his computer to the internet, as well as to the network of the organisation.

The penetration tester asks the receptionist if he could just quickly connect his network cable to the network access point as he urgently needs to check his e-mail. The receptionist knows that it is company policy to assist customers in the reception area and is not aware of the dangers involved in providing access to the network. Since there is no company policy regarding who may and may not

connect to the network access point, it is agreed that the penetration tester may use the network access point. With access gained to the organisation's network, the penetration tester is now able to hack into the inner network and extract sensitive information as needed.

The ethical concern for this scenario is as follows: *Is it ethical to report a social engineering penetration test as successful when the incident occurred because the employee was correctly performing his or her duty?*

4.5.2.4 Penetration test reporting scenario

This scenario deals with the information that is required for the penetration tester's report that is sent back to the organisation.

When an employee is susceptible to a specific penetration test, his or her details can potentially be recorded. The employer may request a detailed report listing the names of all employees who were susceptible to these attacks.

The intent of a penetration test is to help the organisation find the vulnerabilities in their security. If the vulnerability is an employee, management might feel the need to get rid of the employee to reduce the vulnerability [85]. This may not be the right action as the employee may be doing his or her job correctly, or perhaps the employee was not trained to identify social engineering attacks.

The solution to the problem is rather to train the employee to correctly identify social engineering attacks. An employee who already has been subjected to a social engineering penetration test will be much more vigilant than a new employee who has never heard of social engineering before. If management decides to keep this employee, it might be detrimental to his or her future career if the employee's name is recorded in the penetration testing report.

The ethical concern regarding this scenario is as follows: *Is it ethical to provide the names of employees who were susceptible to penetration tests in a report to an authoritative figure even though this may have consequences for the employees?*

The above concludes examples from the social engineering penetration testing environment. The next section deals specifically with scenarios unique to social engineering research. Note that the next as well as the previous section is associated with scenarios that could also occur in social engineering research.

4.5.3 Social Engineering Research

The following scenarios are all framed in the social engineering research environment.

4.5.3.1 Awareness research and debriefing scenario

In this scenario it is required to measure the level of susceptibility of a group of participants to social engineering attacks by using social engineering research. Testing whether a person is susceptible to social engineering requires that social engineering techniques be applied or that social engineering examples be provided to the participant.

When participants are found to be susceptible to these techniques or examples, they may feel as if they were fooled or tricked during the experiment. This can lead to the participants doubting their own decisions and causing them to consider themselves gullible to fall prey to the tactics.

In most of these cases when social engineering awareness was tested, the general consensus was that participants will easily fall prey to these types of social engineering attacks [5, 21, 86]. The attacks tested during such an experiment will determine whether the individuals involved are overly helpful and accommodating. It is common to human nature to act in a helpful and accommodating manner towards other individuals and thus it is not necessarily inappropriate to fall prey to a social engineering attack.

The effect on the participant can be minimised provided that the participant is correctly debriefed. The debriefing session will include a one-on-one discussion between the researcher and the participant. The participant can then be informed and reassured that it is common human nature to fall prey to certain social engineering attacks.

The ethical concern regarding this scenario is as follows: *Is it ethical to conduct social engineering awareness research and how should the participant be debriefed?*

4.5.3.2 Informed consent scenario

This scenario requires participants to provide informed consent for a specific research scenario, but then to be subjected to another social engineering based research scenario. The participant must be fooled into thinking that he or she is part of a different study so that the participants will not be biased against social engineering attacks during the experiment.

In order to receive accurate results from the participants during a social engineering based research experiment, they may not be aware of the type of experiment they are subjected to. The test can be framed to be a normal test so that the participants are unaware that they are partaking in a social engineering research experiment.

The researcher, who requests informed consent from the participant for a different research scenario than the one the participant will eventually be subjected to, is not doing it to be malicious or harmful to the participant. The researcher is merely trying to limit any bias that a participant might have against the specific field and to ensure accurate experimental results from the participant.

The ethical concern regarding this scenario is as follows: *Is it ethical to mislead a participant about informed consent when such consent is required to gain accurate results from the social engineering research experiment?*

4.5.3.3 Sensitive information scenario

In the final scenario, a participant provided informed consent for the specific social engineering based research experiment. However, the researcher goes beyond the information that the participant has provided and gathers additional information to have a better overview of the participant for the sake of the experiment.

The participant who signed up for the social engineering experiment is not aware of the extent to which the researcher will be performing information gathering and social engineering attacks. Since the researcher gathered additional information of which the participant is unaware, the researcher is also unaware of how the participant may react to this additional information.

The information gained from the information gathering experiment might be sensitive or harmful to the participant and it might be unethical of the researcher to reveal this information to the experiment group. The researcher may also misuse the information he has gathered from the participant and appeal to the latter's emotions to manipulate him or her to take part in the experiment.

The ethical concern regarding this scenario is as follows: *Is it ethical during a social engineering research experiment to utilise information about the participant that may be harmful or sensitive to the participant?*

This concludes the section on the ethical concerns regarding social engineering. All those that have been identified in this section will next be discussed from the point of view of the different ethical approaches to normative ethics.

4.6 ETHICAL CONCERNS AND THE CORRESPONDING ETHICAL PERSPECTIVE

This section discusses each of the identified ethical concerns by examining how the different normative approaches can be used to address these concerns. The three different normative approaches that will be used are the virtue ethics perspective, utilitarianism perspective and deontological perspective.

4.6.1 Is it ethical to use social engineering to gain the trust of an individual?

4.6.1.1 Virtue Ethics

When the person performing social engineering is judged, his actions may be seen as negative since he misused his skills set to gain the trust of an innocent victim. For example, the Australian radio presenter did not comply with either the IEEE or the ACM code of ethics, specifically regarding honesty and

trustworthiness. He also did not respect the privacy of the British royal family. From a virtue ethics perspective, such behaviour is unethical.

4.6.1.2 Utilitarianism

In this scenario, the goal of the presenter was to gather information about the royal family for the public's entertainment. The intent was never to harm the nurse, just to gain information. However, in order to gather this information, the presenter used social engineering techniques such as lying and false pretence to mislead the nurse and goad her to give out sensitive information. Thus, social engineering techniques were used to gain the trust of an individual in a deceptive manner. However, when one takes into account the effect on Kate and the royal family, it has to be admitted that since they are public figures, Kate's pregnancy would have been reported in the media in some form or other, sooner or later. The Australian radio presenters were just the first ones to make this information public.

Nevertheless, it is important to note that this is an extreme example as the nurse committed suicide soon after this event. Thus, the verdict would be that had the suicide not happened, the likely utility would have been positive, but if the suicide followed directly from this particular act and its consequences, then the utility was clearly negative. The suicide of the nurse cannot be ignored, as it was the consequence of this specific act and it has to be considered that the same consequences can occur in a similar instance of the ethical question. Hence, from a utilitarianist perspective, this is seen as immoral and thus unethical.

4.6.1.3 Deontology

Deontology is directly concerned with whether a universal world rule was breached to perform the action. For a deontologist, it is unethical to breach any of the universal moral rules, and thus a deontologist is not allowed to lie or manipulate people. To gain trust in this scenario, lying and trickery were used and several moral rules were broken. From a deontological perspective, this is unethical.

4.6.2 Is it ethical when delegated permission is used to perform social engineering techniques for public comical relief?

4.6.2.1 Virtue Ethics

The social engineer received delegated permission to perform the social engineering attack. Even though the social engineer obtained delegated permission, it is not the same as permission from the victim — thus some of the ethical concerns are delegated to the individuals who provided the permission to the social engineer.

The question to ask is whether performing pranks on other people, in other words lying to and intentionally misguiding and misleading them, makes the person performing the prank a more virtuous person. The radio prankster specifically does not avoid injuring the other person's reputation and also does not respect the latter's privacy. Thus, the radio prankster does not comply with the IEEE or the ACM code of ethics and it can be concluded that from a virtue ethics point of view his actions are unethical.

4.6.2.2 Utilitarianism

The majority of the large public audience gained entertainment from this scenario, which outweighs any consequences that the attack may have had on the targeted individual. From a utilitarian perspective, this is ethical as the joy and laughter of the majority outweighs the minor momentary humiliation of the targeted individual.

4.6.2.3 Deontology

Although the social engineer gained delegated permission to perform the social engineering techniques, permission was not granted by the victim self. There was also trickery and lying involved in performing the social engineering attack, which go against the moral rules of deontologists. From a deontological perspective, the radio prankster's action was unethical.

4.6.3 Is it ethical to use information gathering techniques to provide participants with false information and to exploit them for either financial gain or fame?

4.6.3.1 Virtue Ethics

If this scenario was judged by the outside world, they would not see it as good as the con artist or psychic (social engineer) exploited the victim. The social engineer also provided false information to the victims, which gave them false hope, thus showing the social engineer as a bad person. The con artist did not adhere to the IEEE code of ethics and was not honest and realistic in making claims, as the claims are not based on available data. The con artist also did not comply with the ACM code of ethics as he or she provided unsubstantiated information that may cause the victim harm. Thus, his or her actions were unethical from a virtue ethics point of view.

4.6.3.2 Utilitarianism

In this scenario the aim of the con artist was to gain fame and/or fortune through utilising social engineering techniques. The rest of the world does not gain anything from his or her actions. Even if the victim gained some false hope from the act, this false hope did not outweigh the clear wrong-doings by the con artist. The only individual who gained anything was the con artist, whose behaviour was seen as unethical from a utilitarianist perspective.

4.6.3.3 Deontology

The social engineer exploited and lied to the victim, thus breaking or violating several moral rules. From a deontological perspective, this is unethical.

4.6.4 Is it ethical for the employee to bear the consequences of the successful infiltration, when the actual reason for the successful infiltration is not due to the employee's negligence?

4.6.4.1 Virtue Ethics

The question here is whether the employee should bear any consequences when he or she was not negligent. The ethical concern thus focuses on the employer, who is the one initiating the consequences.

Does punishing the innocent employee make the employer a more virtuous person? The employer has to accept responsibility for his or her decisions according to the IEEE code of ethics. The employer would be acting unethically if the employee had to bear the consequences of the employer's decisions. The ACM code of ethics requires the employer to be fair and thus no harm may befall the employee due to the employer's decisions. From a virtue ethics perspective, this is unethical, as the employee should not suffer the consequences when he or she was merely following instructions from the employer.

4.6.4.2 Utilitarianism

The well-being of the employee does not affect the majority of the community. Only the employee is concerned whether there will be any consequences to him or her. If there are any consequences to the employee, these will surely outweigh the consequences to the community, of which there are none. In either case, whether the employee suffers consequences or not, this will still outweigh the consequences to the community. Thus, the harm done to the employee due to the successful infiltration will always be unethical, no matter whether there are consequences to the employee or not. From a utilitarian perspective, such behaviour is unethical.

4.6.4.3 Deontology

The employee was not negligent and thus does not deserve to suffer any consequences. It is not the employee who is at fault, provided that the employee followed the correct instructions from his or her

superior. Any repercussions or consequences for the employee due to the successful infiltration will therefore be unethical from a deontological perspective.

4.6.5 Is it ethical to exploit a personal weakness of an employee when it is known to be common human nature to fall prey to this type of attack?

4.6.5.1 Virtue Ethics

It is known to be common human nature to fall for this type of attack. In this specific scenario the attack involves bribing the individual with an offer that is unreachable under normal circumstances. Since the social engineer takes advantage of an already known weakness, this does not make the social engineer a more virtuous person. Moreover, the IEEE code of ethics clearly states to reject bribery in all its forms. The social engineer provides a bribe to the victim and this is unethical according to the IEEE code of ethics. The social engineer is also not acting in an honest and trustworthy manner, which violates the ACM code of ethics. From a virtue ethics perspective, such behaviour is unethical.

4.6.5.2 Utilitarianism

In the specific penetration testing scenario the employee will be reassured that this is a common human vulnerability and thus the employee will be more vigilant and alerted to this type of attack in the future. If the employee is not unfairly dismissed, this may have further benefit for the organisation as the employee can use the opportunity to educate the rest of the staff to be more vigilant. The employee who was vulnerable to the attack can have a huge positive impact on his organisation by warning others against such an attack. From a utilitarian perspective, the harm done to the employee does not weigh up to the clear advantage he or she might now provide to the organisation through educating other personnel. Due to the eventual huge benefit to the organisation, this action is seen as ethical from a utilitarian perspective.

4.6.5.3 Deontology

The social engineer misused a common human vulnerability to exploit the victim. Exploiting other humans for personal gain clearly defies several rules of morality. From a deontological perspective, this is unethical as the action broke several rules of morality by tricking and exploiting the employee.

4.6.6 Is it ethical to report a social engineering penetration test as successful when the incident occurred because the employee was correctly performing his or her duty?

4.6.6.1 Virtue Ethics

The focus is not on the employee who performed wrongly, but rather on the guidelines and regulations of the organisation that need to be addressed and corrected. The employee did what he thought was required, thus showing good character in accordance with virtue ethics. The guidelines and regulations were followed correctly by the employee and thus the fault lies with the guidelines and regulations, not with the employee. Even though the guidelines and regulations were misused, the employee acted virtuously. Reporting the social engineering penetration test as successful will not harm the employee for performing his or her duties correctly as it will lead to the guidelines and regulations being corrected. The IEEE code of ethics requires the social engineer to be honest and realistic based on available data. Hence it would be ethical for the social engineer to report the penetration test as successful. Similarly, the ACM code of ethics requires the social engineer to report on his findings honestly while avoiding harm to others. From a virtue ethics perspective, such behaviour is therefore ethical.

4.6.6.2 Utilitarianism

Although reporting on the successful penetration test can cause the employee to suffer consequences, the organisation greatly benefits from it. By seeing the report, the organisation can better their guidelines and regulations so that future penetration tests as well as real attacks will not be successful. From a utilitarian perspective, the benefits of the majority of the organisation outweigh the possible

consequences on the employee. Reporting on the successful penetration test is seen as ethical from a utilitarian perspective due to the benefit of the organisation.

4.6.6.3 Deontology

The outcome of the penetration test is that it was successful. From a deontological perspective it is required to report on the social engineering penetration test and to confirm that the information in the social engineering penetration test report is correct and accurate. In order to oblige to the rules and not be dishonest about the facts, the social engineering penetration test should be reported and accurately so. From a deontological perspective, it is ethical to report the social engineering penetration test as successful.

4.6.7 Is it ethical to provide the names of employees who were susceptible to penetration tests in a report to an authoritative figure even though this may have consequences for the employees?

4.6.7.1 Virtue Ethics

The harm to the employee must be pre-empted by the social engineer penetration tester according to the IEEE and ACM code of ethics. The penetration tester should inform the authoritative figure of the correct way to assist the employees and not to harm or punish them for their mistakes. Furthermore, both ethical codes also require the social engineer to report honestly and correctly on all their findings. Since it is required of the social engineer to pre-empt all harm and report honestly and accurately, it is deemed ethical for the social engineer to report fully on all findings, including the names of the employees involved. From a virtue ethics perspective, this is ethical.

4.6.7.2 Utilitarianism

In this example, it is important to note that the good of society or the good of humankind is the ultimate utilitarian consideration. In the case where training the employee and benefit to the organisation outweigh the dismissal of the employee and benefit to the organisation, the first option constitutes the

morally obligatory choice. Since this benefits the majority of the organisation, it is ethical according to the utilitarianist perspective.

4.6.7.3 Deontology

If it is assumed that the penetration tester is required by management to report the full detail of the penetration test to the organisation, it would be ethically correct to disclose the names of vulnerable employees as the focus is on the rule stating that the penetration tester should provide a report with full details.

From a deontological perspective, such behaviour is ethical as the penetration testing follows the moral rule of full disclosure.

4.6.8 Is it ethical to conduct social engineering awareness research and how should the participant be debriefed?

4.6.8.1 Virtue Ethics

From an outside perspective it would be seen as ethical as it is required in research to perform social engineering awareness testing. However, it will only be seen as good if sufficient time is spent on debriefing each of the participants.

The IEEE code of ethics requires the researcher to accept responsibility in making decisions and to avoid harm to others. The ACM code of ethics requires the researcher to contribute to society and human well-being whilst also avoiding all harm to others. Avoiding harm to others entails the adequate debriefing of all participants.

From a virtue ethics perspective, such behaviour is seen as ethical, provided that the researcher ensures that the participant is correctly debriefed to the best of the researcher's ability.

4.6.8.2 Utilitarianism

Research is always needed even if such research may be harmful to some of the participants. As long as the ultimate goal of the research is to improve society as a whole, it will be seen as ethical from a utilitarian perspective. From such a perspective, any research that is performed to better the greater whole of society is seen as ethical.

4.6.8.3 Deontology

Social engineering awareness testing has the ultimate goal to pose questions to the participants to determine whether they are susceptible to social engineering. These questions are developed in a way to trick the participant and can thus cause the participant to answer the question in a manner that shows susceptibility. As deontology has a rule that participants should not be lied to or tricked during research, such behaviour will be seen as unethical.

4.6.9 Is it ethical to mislead a participant about informed consent when such consent is required to gain accurate results from the social engineering research experiment?

4.6.9.1 Virtue Ethics

Participants are required to give informed consent for the particular research experiment in which they are about to participate. The specific participant might not have given informed consent for an experiment that involves social engineering research. If the participant had been aware of the fact that he or she would be taking part in a social engineering experiment, he or she might have chosen to not form part of the research experiment.

Both the IEEE and ACM codes of ethics require the researcher to be honest with the research participants and thus it is unethical to be dishonest about the informed consent. Informed consent cannot be given by a participant if he or she is being misled.

From a virtue ethics perspective, the participant must be fully aware of the research experiment that he or she is going to be part of. Since informed consent is not given for the social engineering research experiment, it is considered unethical to trick the participant to be part of such experiment.

4.6.9.2 Utilitarianism

The social engineering research experiment may be harmful to a participant who gave informed consent for a different research experiment. In some scenarios, accurate results can only be gained if the participant is unaware of partaking in the research. Participants may behave differently if they know that they are part of a social engineering research experiment. It is important to limit the bias that the participant would have towards such an experiment to ensure the most accurate results from it.

However, since accurate results are required to improve society as a whole and since they outweigh the harm that can possibly be done to the participant, this action is regarded as ethical from a utilitarian perspective.

4.6.9.3 Deontology

The participant did not give informed consent for participating in the social engineering research experiment. Since this implies violation of one of the social engineering research rules, such action can already be seen as unethical. The participant may also feel bad for being tricked since he or she was not aware of participating in a social engineering research experiment.

From a deontological perspective, such behaviour is unethical as the participant is fooled into participating in a research experiment that he or she did not sign up for.

4.6.10 Is it ethical during a social engineering research experiment to utilise information about the participant that may be harmful or sensitive to the participant?

4.6.10.1 Virtue Ethics

Such information may harm the participant unbeknown to the researcher. The researcher may harm the participant by revealing sensitive information about the participant without intending to do any harm. Both the IEEE and ACM codes of ethics specifically state that the researcher may not cause harm to any of the participants. From a virtue ethics perspective, such action is seen as unethical since the researcher may potentially harm the participant.

4.6.10.2 Utilitarianism

Research is needed even if the information and method used to perform the social engineering research might be harmful to some of the participants. In this example it is important to note that the good of society or the good of humankind is the final measure or consideration in a utilitarianist approach. Hence, from a utilitarian perspective the research will still be seen as ethical, since the obligatory goal of all research is to improve society as a whole, and this commendable goal outweighs any negative consequences to the participant. The researcher can only perform the act that will most optimally improve the society, since (according to the utilitarianist perspective) the best available solution is considered the morally obligatory choice.

4.6.10.3 Deontology

The participant did not provide informed consent to the researcher to utilise his or her harmful or sensitive personal information during the social engineering research experiment. Since informed consent is required, especially when dealing with harmful or sensitive information, the deontological rules were not followed. From a deontological perspective, such behaviour is unethical.

To summarise this section, tables 4.1, 4.2 and 4.3 list all of the ethical concerns in the three environments and whether they are ethical from the point of view of each of the different ethical perspectives.

Table 4.1: Ethical Concerns in Public Communication

	Virtue ethics	Utilitarianism	Deontology
Is it ethical to use social engineering to gain the trust of an individual?	No	No	No
Is it ethical when delegated permission is used to perform social engineering techniques for public comical relief?	No	Yes	No
Is it ethical to use information gathering techniques to provide participants with false information and to exploit them for either financial gain or fame?	No	No	No

Table 4.2: Ethical Concerns in Penetration Testing

	Virtue ethics	Utilitarianism	Deontology
Is it ethical for the employee to bear the consequences of the successful infiltration, when the actual reason for the successful infiltration is not due to the employee's negligence?	No	No	No
Is it ethical to exploit a personal weakness of an employee when it is known to be common human nature to fall prey to this type of attack?	No	Yes	No
Is it ethical to report a social engineering penetration test as successful when the incident occurred because the employee was correctly performing his or her duty?	Yes	Yes	Yes
Is it ethical to provide the names of employees who were susceptible to penetration tests in a report to an authoritative figure even though this may have consequences for the employees?	Yes	Yes	Yes

Table 4.3: Ethical Concerns in Social Engineering Research

	Virtue ethics	Utilitarianism	Deontology
Is it ethical to conduct social engineering awareness research and how should the participant be debriefed?	Yes	Yes	No
Is it ethical to mislead a participant about informed consent when such consent is required to gain accurate results from the social engineering research experiment?	No	Yes	No
Is it ethical during a social engineering research experiment to utilise information about the participant that may be harmful or sensitive to the participant?	No	Yes	No

The following section provides practical examples on how this research can be utilised.

4.7 PRACTICAL EXAMPLES WITH REGARD TO THE ETHICAL CONCERNS

This thesis has now provided the reader with ten different ethical concerns and how to reason about these ethical concerns from the different ethical perspectives. Three practical examples of where this research can be utilised are suggested next, such as in ethical committees, for teaching ethics in computer security and as an ethical guideline for penetration testers.

4.7.1 Ethical committees

Ethical committees perform a tedious job which entails verifying that all research performed adheres to several ethical guidelines. For the social engineering field specifically there is no formalised set of rules for measuring the ethical impact of a social engineering attack.

The current research can be used by ethical committees as a tool to measure the ethical impact of social engineering based research. It provides ethical committees with what each of the three different ethical perspectives have to say about each of the different ethical concerns. This research can also be used

to answer specific ethical questions and to determine whether a single action in social engineering is ethical or not.

For example, a student approaches the ethical committee and wants to conduct social engineering based research that is specific to a certain organisation. He also wants to research the effects that social engineering attacks may have on the structure of the organisation. The table of ethical concerns (Table 4.3) allows one to easily determine the major ethical concerns are associated with this research. It also provides both the student and the ethical committee an easier way to measure the ethical viability of the research proposal.

This research can also provide an ethical committee with the three different ethical perspectives and how they are addressed in terms of social engineering. From tables 4.1, 4.2 and 4.3 one can clearly see that when the ethical committee examines project proposals based on a utilitarianism perspective, more projects will be approved than when the same project proposals were to be examined from a deontological perspective.

4.7.2 Teaching ethics in computer security

The research can also be utilised to teach computer science students the ethical impact of social engineering in the field of computer security. As social engineering is entrenched in both computer science and social psychology, it is important for computer scientists to understand pertinent ethical concerns when dealing with individuals.

The table of ethical concerns can also be utilised to teach students the difference between the three different normative ethical approaches and how the reasoning of each of the ethical perspectives can be utilised to determine whether an action is ethical or not. It is furthermore important for students to understand the difference between the ethical perspectives and how the ethical measurement of each of these perspectives differs.

Since Table 4.1 and Table 4.2 provide practical examples on how to judge whether a social engineering action is ethical or not, it can be expanded for students to focus on other fields within the domains of computer security and computer science.

4.7.3 Ethical guideline for penetration testers

Penetration testers often have to decide whether a certain penetration test would be deemed ethical or not [85]. Also, among the scenarios provided, there were more that could be taken directly from the penetration testing environment as it is such a difficult environment in which to judge whether a certain action is ethical or not.

The present research provides penetration testers with a good guideline for measuring their applicable social engineering penetration tests. Table 4.2 can assist the penetration testers when it comes to ethical concerns about reporting on a certain successful infiltration. It is important to them to report their information in an ethical manner as the outcome potentially has a major impact on an employee's life (e.g. if the employee is dismissed due to the results of the penetration testing report).

Penetration testers will also benefit from having available the different ethical perspectives on each of the different ethical concerns. Being informed about the different ethical perspectives allows the penetration tester to examine the ethical concerns with the different perspectives and to make an informed decision about their actions.

The thesis concludes the researcher's work by providing a summary of the ethical concerns about social engineering, and how this research can be utilised in practice.

4.8 CONCLUSION

Social engineering is deeply entrenched in the fields of computer science and social psychology. Knowledge of both of these disciplines is required to apply social engineering based techniques. Since all of these techniques are ordinarily performed on human participants, the ethical impact of social engineering on these participants needs to be considered. Several ethical concerns and requirements need to be taken into account when social engineering research is conducted to ensure that no harm comes to the participants.

The problem is that these requirements have not yet been formalised and most researchers are unaware of the ethical concerns related to social engineering research. This thesis firstly addresses this problem

by providing the reader with a thorough background of the three main perspectives derived from the normative ethics approach.

The thesis secondly discusses three environments in which social engineering can occur, i.e. public communication, penetration testing and social engineering research. As the social engineering research environment is such a broad field, it can contain scenarios from both public communication and penetration testing. Each of the three environments is subdivided into several different and applicable scenarios.

These scenarios are used to develop and provide frames in which the ethical concerns regarding social engineering were proposed. Each scenario is associated with a single ethical concern, while each ethical concern has a scenario in which to frame the ethical concern to test whether the action taken is ethical or not.

The ethical concerns that are proposed are measured against each of the three different ethical perspectives, namely virtue ethics, utilitarianism and deontology. Each ethical concern is addressed by utilising all of the ethical perspectives. This thesis furthermore provides practical examples of where this research can be used, for instance as a tool for ethical committees, to teach ethics in computer security, and as an ethical guideline for penetration testers.

The purpose of this chapter is to inform the reader of the ethical concerns that had to be adhered to throughout this thesis. The research into ethical concerns clearly depicted that there is a very high risk in causing harm to individuals whilst performing actual social engineering attacks during research. This had a major impact on this thesis and caused the final proposed SEADM to be verified using mathematical proofs and attack examples instead of real world based testing. Great effort has been taken to ensure that the mathematical proofs and attack examples accurately simulate real world based testing and each of these techniques has their own chapter dedicated to them.

This chapter now concludes all the literature that needs to be taken into consideration throughout this thesis. The literature chapters ensured that the following items were addressed:

- A singular definition, used throughout the thesis, for social engineering has been defined.

- The intial iteration of the SEADM, which had originally initiated this thesis, has been introduced to the reader.

- The ethical concerns that had to be taken into consideration throughout the thesis and the justification as to why the model was verified using mathematical proofs and simulated social engineering attack examples.

This now concludes all of the required literature that the reader needs to be aware of and the rest of the thesis contains the main contribution sections. The first contribution chapter, Chapter 5, focuses on addressing the problem that there is no formal social engineering attack framework defined prior to this thesis. The following chapter attempts to rectify this by proposing a formalised social engineering attack framework.

CHAPTER 5 SOCIAL ENGINEERING ATTACK FRAMEWORK

5.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the proposed Social Engineering Attack Framework (SEAF). This chapter is based on a conference paper entitled "Social Engineering Attack Framework" which has previously been published by the author at the Information Security for South Africa conference in 2014 [87]. This conference paper proposed the SEAF and detailed the uses of such a framework. The purpose of the SEAF, in terms of this thesis, is to have a framework which can be utilised to formally develop social engineering attack examples. This chapter is included here as the first contribution chapter of this thesis as it provides the reader the basis on how social engineering attack examples were derived and formalised. This chapter utilises the ontological model which was provided in Chapter 2 as a baseline for the development of the SEAF.

5.2 INTRODUCTION

The proposed ontological model, in Section 2.6, includes components of a social engineering attack and divides the attack into different classes and subclasses. The two classes of a social engineering attack are: Direct communication and indirect communication. The direct communication class is further divided into two subclasses: Bidirectional communication and unidirectional communication. A social engineering attack is then further explained to contain the following components: one Social Engineer; one Target; one or more Compliance Principles; one or more Techniques; one Medium; and one Goal [9].

Although the ontological model contains all the components of a social engineering attack, an ontological model struggles to depict temporal data, such as flow and time [55]. One of the main features of an ontology is that it separates the domain knowledge from the operational knowledge [55]. Due to this shortcoming, the ontological model is not sufficient to depict the process and the steps involved in executing a social engineering attack. The purpose of this chapter is to present a social engineering attack framework which, in conjunction with the ontological model, investigate the attack process in detail. The framework refers to the components in the ontological model, but focuses on the process flow starting at the point at which an attacker initially thinks about gaining sensitive information from some target up to the point of succeeding in the goal of gaining this information.

The ontological model provides the basic structure of a social engineering attack whereas the social engineering attack framework adds both time and flow components. The combination of the ontological model and the attack framework can be used to generate social engineering attack scenarios and to map historical social engineering attacks to a standardised format. These scenarios are useful to educate individuals about social engineering and to gauge their awareness of social engineering. Scenario generation is also useful in the development of countermeasures against attacks. Having a standardised formulation of a social engineering attack as well as the flow and time events, allow researchers to compare different social engineering attacks.

The remainder of the chapter is constructed as follows. Section 5.3 provides a background on social engineering attacks and further expandes on the ontological model that was proposed in Chapter 2. Section 5.4 discusses the proposed social engineering attack framework and section 5.5 provides some applications of the social engineering attack framework. Section 5.6 concludes this chapter.

5.3 DEFINING SOCIAL ENGINEERING ATTACKS

There are many models and taxonomies concerning social engineering attacks which are explored and analysed in the Chapter 2. The most prominent attack model in the field of social engineering is Kevin Mitnick's social engineering attack cycle as described in his book, *The Art of Deception: Controlling the Human Element of Security*, [4]. Mitnick's attack model has four phases: research, developing rapport and trust, exploiting trust and utilising information. These four phases are not explained in great detail in Mitnick's book.

The following picture is a representation of Mitnick's attack cycle. Figure 5.1 depicts the four phases and the flow between each of the phases. Each of these phases are briefly discussed below as explained in Mitnick's book.

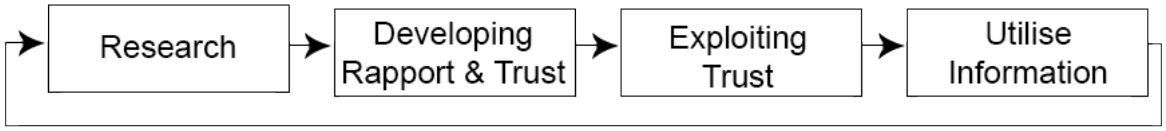

Figure 5.1: Kevin Mitnick's Social Engineering Attack Cycle

Research is an information gathering process where information about the target is retrieved. The attacker should know as much as possible about the target before starting the attack.

The next phase is the **Development of the rapport and trust** with the target. A target is more likely to divulge requested information to an attacker if he trusts the attacker. According to Mitnick [4], rapport and trust development can be done by using insider information, misrepresenting an identity, citing those known to the victim, showing a need for assistance, or occupying an authoritative role.

When a target appears to trust an attacker, the attacker **Exploits the trust** to elicit information from the target: this can either take the form of a request for information, a request for a specified action from the victim or, alternatively, to manipulate the victim into asking the attacker for help [4]. This phase is where the previously established relationship is abused to get the initially desired information or action.

Finally, the outcome of the previous phase is **Utilised** to reach the goal of the attack or to move on to further steps which may be required to reach the goal.

A trivial example is when an attacker supposedly needs to connect to an organisation's network. As a result of his research the attacker finds out that a help-desk staff member knows the password to the organisation's wireless network. In addition, the attacker found personal information regarding the staff member who has been identified as the target. The attacker initiates a conversation with the target, using the acquired information to establish trust; in this case the attacker misrepresents himself as an old school acquaintance of the target. The attacker subsequently exploits the established trust by asking permission to use the company's wireless network facility to send an e-mail. The help-desk

attendant is willing to supply the required password to the attacker due to the misrepresentation, and is able to gain access to the organisation's network and achieve his objective.

The ontological model, as discussed and depicted in Chapter 2 in Figure 2.5, defines that 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. The attack can be split into more than one attack phase, each phase handled as a new attack according to the ontological model.

Direct communication, where two or more people communicating directly with each other, is subdivided into "Bidirectional communication" and "Unidirectional communication". Bidirectional communication occurs when both parties participate in the conversation. For example, an e-mail is sent from the attacker to the target and the target replies to the attacker. Unidirectional communication occurs when the conversation is one-way only: from the attacker to the target. For example, if the attacker sends a message through paper mail without a return address, the target cannot reply to the message. Phishing attacks are also a popular type of attack in this category.

Indirect communication is when there is no actual interaction between the target and the attacker; communication occurs through some third party medium. An example of this type of communication is when the attacker infects a flash drive and leaves it somewhere to be found by some target. The target is curious to find out what is on the flash drive for personal gain or, motivated by ethical consideration, to attempt to find the owner of the flash drive. The target inserts the flash drive into their computer, and the infection on the flash drive is activated.

The ontological model further contains several components as mentioned in the introduction. The goal can be financial gain, unauthorised access or service disruption. The medium is a way of communication such as e-mail, face to face, telephone etc. The social engineer can be either an individual or a group of individuals. The target can either be an individual or an organisation.

Compliance principles refer to the reasons why a target complies with the attacker's request, and techniques include those used to perform social engineering attacks. Examples of techniques include phishing, pretexting, baiting and *quid pro quo* [88]. Examples of compliance principles include [88]:

- *Friendship or liking*: People are more willing to comply with requests from friends or people they like.

- *Commitment or consistency*: Once committed to something, people are more willing to comply with requests consistent with this position.

- *Scarcity*: People are more willing to comply to requests that are scarce or decreasing in availability.

- *Reciprocity*: People are more willing to comply with a request if the requester has treated them favourably in the past.

- *Social Validation*: People are more willing to comply to a request if it is seen as the socially correct thing to do.

- *Authority*: People comply easily to requests given by people with more authority than they have.

Once the compliance principles, techniques and medium have been selected, the attack vector can be set-up and the social engineer can continue to the actual attacking phase.

The next section introduces the proposed social engineering attack framework.

5.4 SOCIAL ENGINEERING ATTACK FRAMEWORK

This section proposes an extension of Kevin Mitnick's original social engineering attack cycle [4]. Mitnick's attack cycle is explained very briefly in his book and does not contain a lot of detail [4]. Mitnick's attack cycle is very broad and is open to interpretation in some aspects. Figure 5.2 depicts the new proposed social engineering attack framework. This framework clarifies Mitnick's phases and is more detailed.

In Mitnick's first phase, the research phase, he states that when executing a social engineering attack one needs to get the most possible information about the target. Even though this is true, this is a very

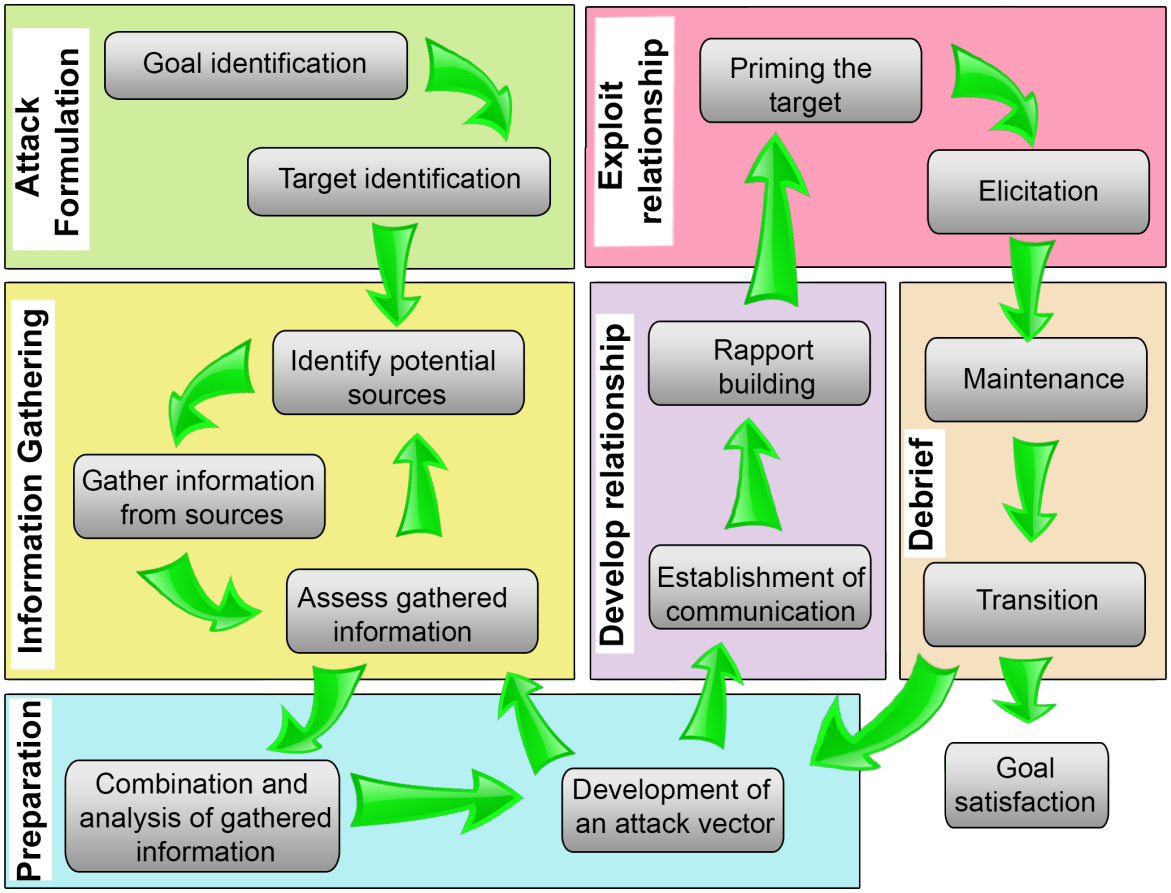

Figure 5.2: Social Engineering Attack Framework

broad statement and it also assumes that the target is already known and that the goal is already set. The author proposes an additional step before gathering the information which is meant for determining what the goal of the attack is and the best possible target to assist with reaching the goal. Once the goal and target is known, the actual information gathering can start. This process is also described in more detail than Mitnick's model as one needs to identify sources of information before anything can be gathered and it is beneficial to assess the gathered information to ensure that there is sufficient information to execute the attack. Mitnick's attack cycle does not contain a preparation phase which is also needed during the social engineering attack. The author proposes a preparation phase which is used to prepare the gathered information and to develop the attack vector that will be used during the social engineering attack.

Mitnick's next phase, development of rapport and trust, is very similar in the proposed framework but the starting point is modelled as a separate step. Establishment of communication is a requirement for any relationship to be built with the target. The gathered information is used to assist in establishing

communication. Once the attacker and the target are communicating, the rapport and trust building can commence.

The third phase, according to Mitnick, is the exploitation phase. This phase also requires more detail than that given in Mitnick's attack cycle. Exploiting a relationship is done with different manipulation techniques and in order for these techniques to work the target has to be in an emotional state where the exploitation is possible. This differs between all human beings and it is thus necessary to first determine what that emotional state is of the target and then get the target into the desired emotional state. Once the target is in the right emotional state, the information can be elicited. The other important step not mentioned in Mitnick's attack cycle is the debriefing step. The target has to be brought back to a normal emotional state to avoid further consequences. The idea is to have the target feel good about giving out unauthorised information instead of feeling guilty about it.

Finally Mitnick has a fourth phase, utilising the information, which the author argues to be not part of the actual social engineering attack. The social engineering attack focuses on attacking the human aspect with the intention to achieve a specified goal, in this case to gain privileged information. This information can be used to perform a different action, but this is no longer part of the social engineering attack. For instance if the information is a password to the system, gaining the password from a person is a social engineering attack whereas using the password to break into the system has no human element to it and is thus not a social engineering attack.

The framework is completed by having a transition phase after debriefing to either go back and gather more information if it is found that more information is needed to be able to complete the attack, or go to the goal satisfaction. Mitnick also states that previous steps can be repeated if the goal is not satisfied, though this is not described in much detail. The proposed framework provides a more precise transition phase specifying the exact phase to return to and repeat if necessary.

The following subsections describe each of these phases in more detail.

5.4.1 Attack Formulation

The first step of a social engineering attack is to address the question "What does the social engineer want?". This goal of the social engineer is the purpose of the entire attack and should be very clear. Once the goal is identified, the target should be selected, as depicted by Figure 5.3. The target can be an individual or a group of individuals.

The target may belong to an organisation that is under attack as part of the goal. For example, the goal may be to infiltrate an organisation and the target is a security guard who possesses information required to accomplish the goal. Both the organisation and the selected target are important in the information gathering phase.

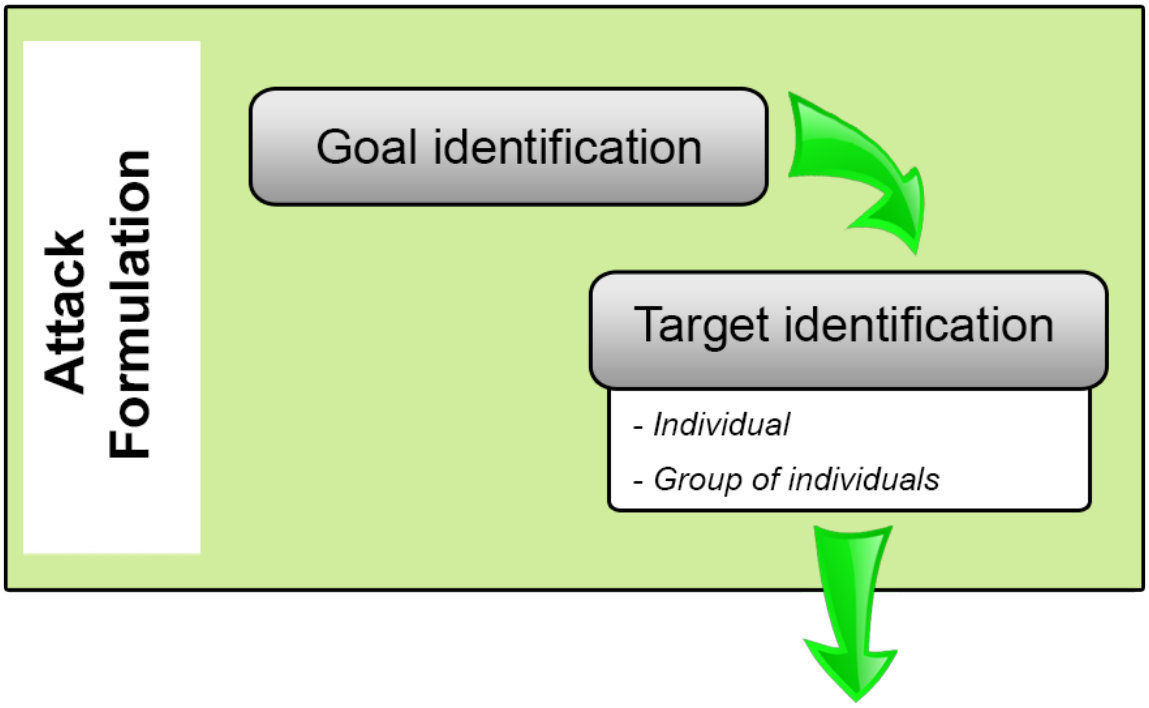

Figure 5.3: Social Engineering Attack: Attack Formulation

5.4.2 Information Gathering

Information gathering is a very important part of the social engineering attack because the probability of developing a trusting relationship with a target is increased by the quality of the information regarding

the target. A target is more likely to share information with the attacker if a relationship exists between the two.

Information is gathered about the target and everything related to the attack. As depicted in Figure 5.4, the first step of gathering information is to 'identify the possible sources' from which information can be obtained. The sources can be anything or anyone with access to the information required for the attack. These sources can be any publicly available sources such as company websites, social networking sites or personal blogs and forums, or private information that is not publicly available. Techniques such as dumpster diving can be used where discarded items are scanned for private information, such as an address on a bank statement. Dumpster diving is the technique of sifting through trash such as medical records or bank statements to find anything that can be useful to the dumpster diver [89].

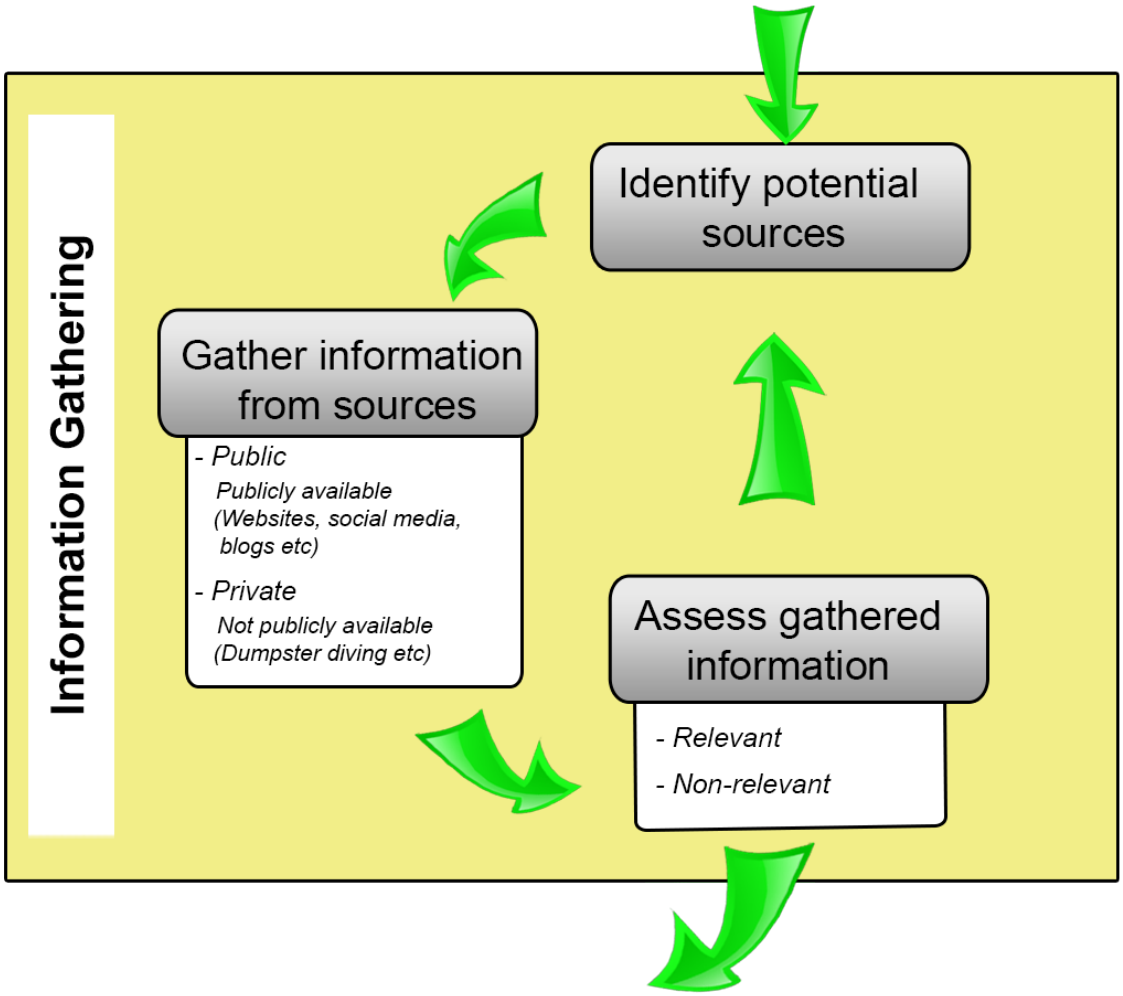

Figure 5.4: Social Engineering Attack: Information Gathering

After gathering information, the information is assessed to be relevant or not. If the social engineer still does not have enough information, he can go back to identifying more sources and restart the information process.

The 'information gathering' phase is repeated until the social engineer is satisfied that sufficient information has been obtained, such that he can start his preparation for the attack.

5.4.3 Preparation

During preparation the social engineer ensures that everything is ready before starting the actual attack. As depicted by Figure 5.5, the first step of this phase is to combine all information gathered to form a bigger picture about the planned attack.

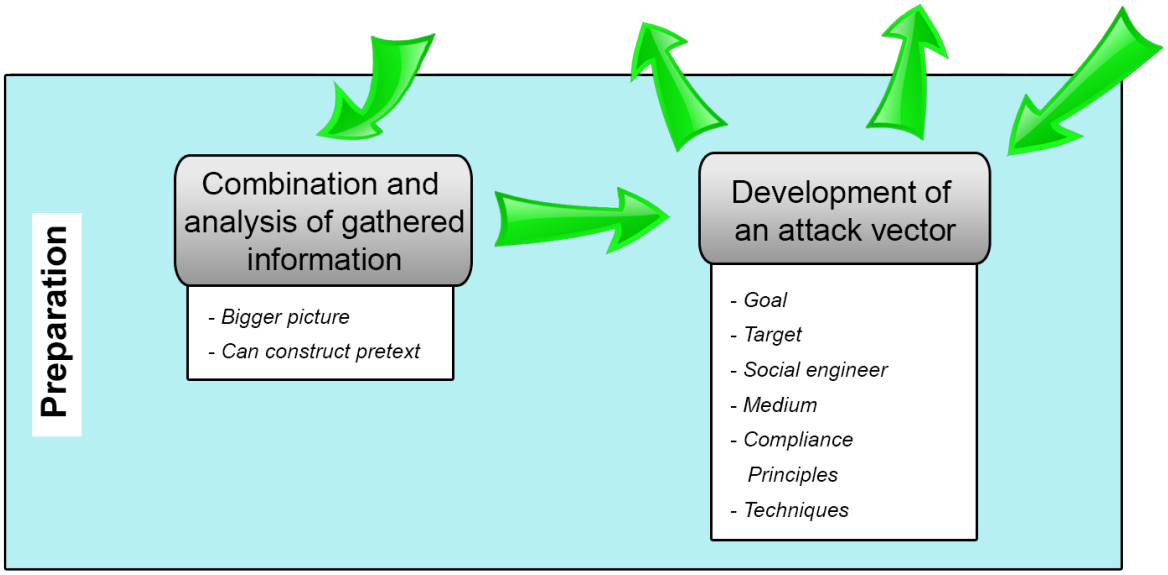

Figure 5.5: Social Engineering Attack: Preparation

This combined view of the scenario can be used for pretexting where a scenario is devised to lure the target into a required action. An effective pretext should be believable and withstand scrutiny from the target. It often relies on the quality of the information gathered on the target's personality. An attack vector is now developed; it should contain all the elements of a social engineering attack [9]. The attack vector is the attack plan which leads to the satisfaction of the goal. It has a goal, a target and a social engineer. In addition, the plan must identify a medium, compliance principles and techniques.

5.4.4 Develop a Relationship

As mentioned previously, developing a good relationship with the target is an essential part of the social engineering attack. If trust cannot be established, the required information is unlikely to be elicited from the target. Figure 5.6 depicts the first step involved in building a relationship with the target, namely the 'establishment of communication' step. This step is executed by using the medium identified during the preparation phase. If a pretext has been included in the plan, it is used along with the initial communication.

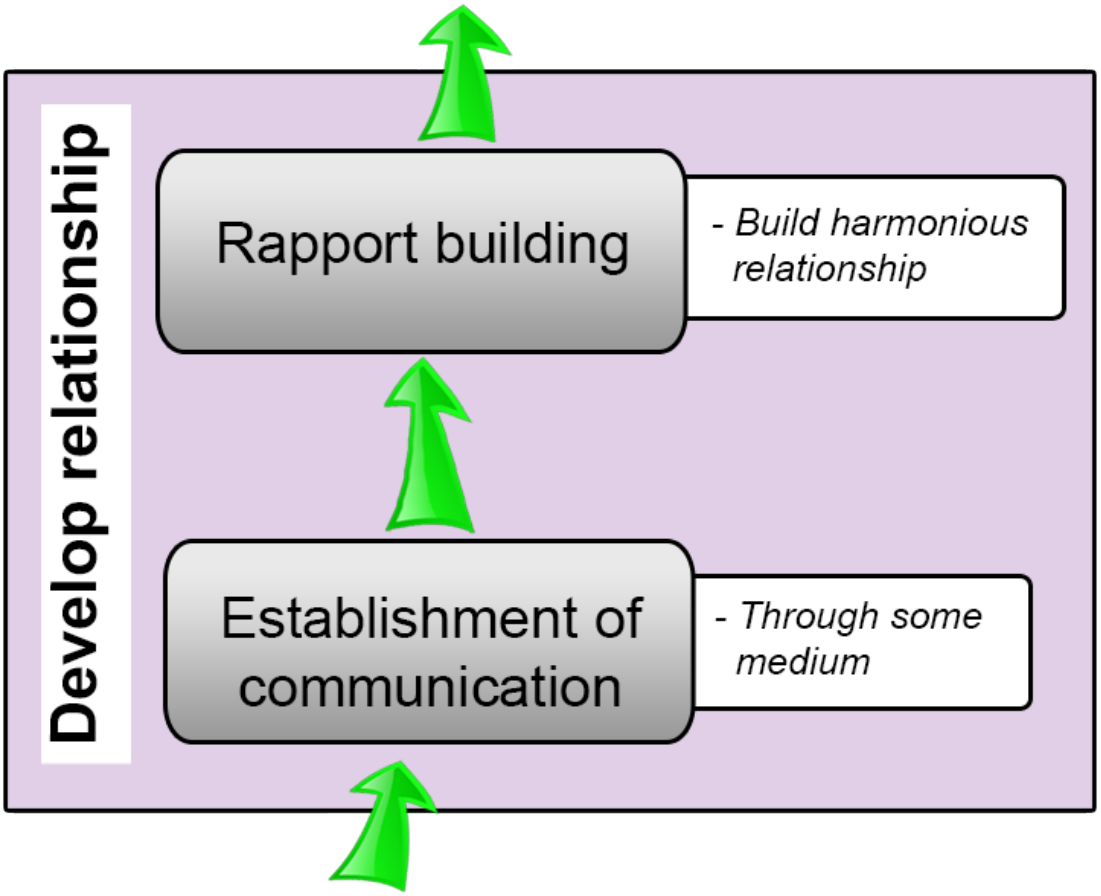

Figure 5.6: Social Engineering Attack: Develop Relationship

The next step in developing a relationship is the 'rapport building'. This entails the actual building of the relationship and establishment of trust using the devised plan. Various techniques can be employed to establish trust. This step is not trivial and can be time consuming. A good pretext simplifies this step. Once the social engineer has built a good relationship with the target, the relationship can be exploited to obtain the information the social engineer requires from the target.

5.4.5 Exploit the Relationship

As depicted in Figure 5.7, exploiting the relationship consists of two parts: 'priming the target' and 'elicitation'. The first part is for the attacker to use manipulation tactics and his preparation to get the target in a desired emotional state suited to the plan, such as feeling sad or happy. For example, relating to a sad story can evoke the target into remembering a sad incident, and subsequently to feel sad.

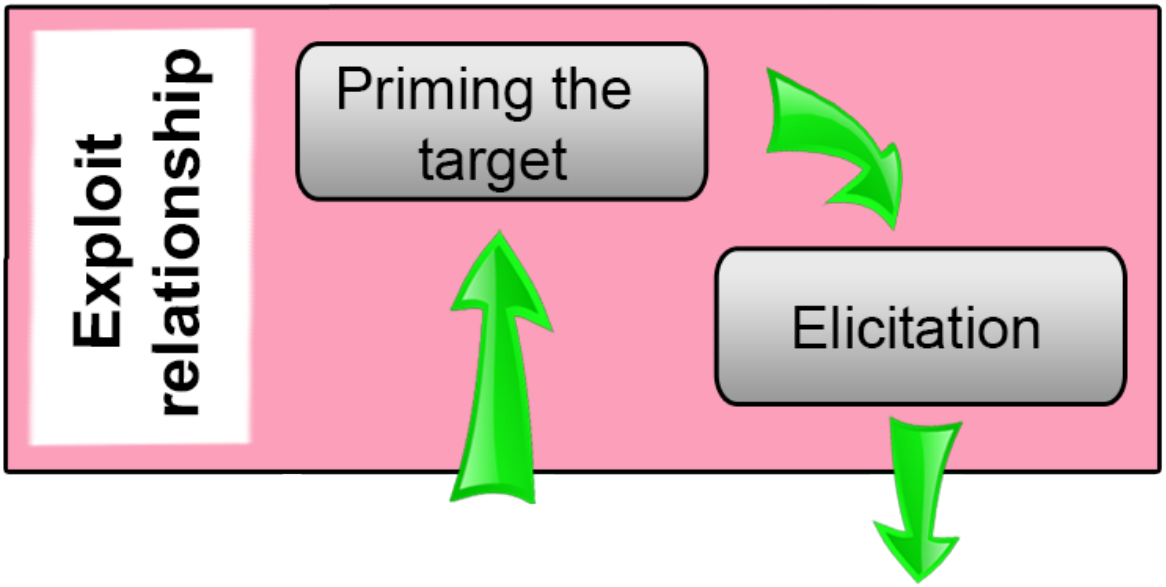

Figure 5.7: Social Engineering Attack: Exploit Relationship

Once the target is in the desired emotional state, the elicitation process can start. At the conclusion of the elicitation phase the social engineer should have obtained the required information from the target. This may be a password which is needed for the eventual satisfaction of the goal of the social engineering attack. After the exploitation phase, it is important to debrief the target.

5.4.6 Debrief

Debriefing the target involves returning the target to a desired emotional state of mind, as shown in the 'maintenance' step in Figure 5.8. It is important for the target not to feel that he was under attack; if he is in a normal state of mind, he will probably not reflect too much on the activities that occurred. For example, if the target had been manipulated into a sad emotional state and the attacker then elicited

a password from him, the target may feel inadequate because he has released sensitive information. This feeling of inadequacy may consequently lead to emotional states such as depression. It may even lead to suicide by the target as evidenced in an incident in 2012 involving the solicitation of private information concerning the British Royal family [46, 67]. During the confinement of Princess Catherine, an Australian radio talk show host socially engineered a staff member of the maternity ward where the princess was a patient, to release information regarding the Princess' condition.

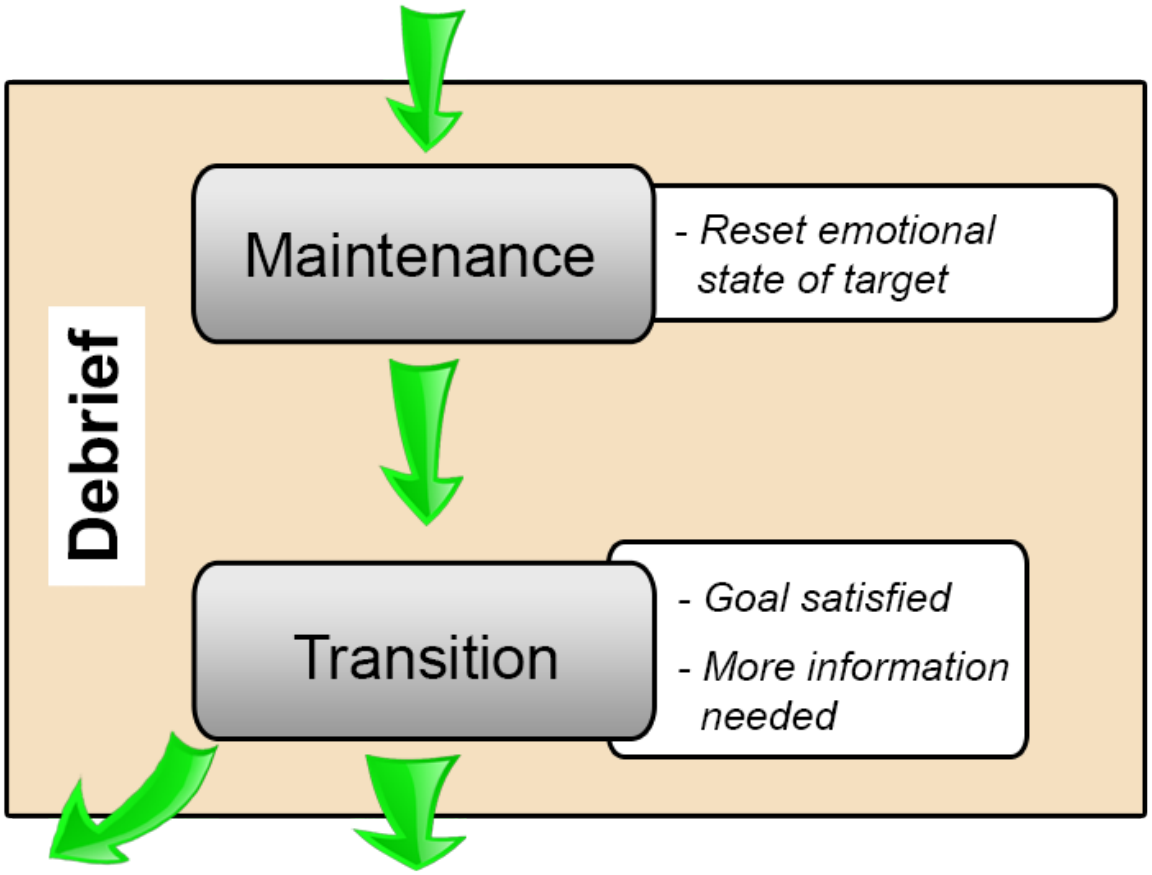

Figure 5.8: Social Engineering Attack: Debrief

Figure 5.8 depicts the next step in the debriefing phase, namely 'transition'. This is where the social engineer either decides that the goal has been satisfied or that more information is needed and the engineer returns to the information gathering phase.

The next section discusses the applications of the framework on two examples.

5.5 FRAMEWORK APPLICATION

This section discusses two examples of well-known social engineering attacks which have also been documented in news articles. Each of the examples are individually mapped to the proposed Social Engineering Attack Framework. This exercise shows that the Social Engineering Attack Framework can be utilised to convert historical social engineering attacks to a standardised format. Having historical social engineering attacks in a standardised format allows one to perform comparisons between two different social engineering attacks. These standardised social engineering attack scenarios can also be used for social engineering training and awareness testing.

The next subsections analyse each of the examples according to the attack framework. Important features of each social engineering attack are mapped to the components in the definition of a social engineering attack. The different components of a social engineering attack are: the type of communication, the social engineer, the target, a medium, a goal, one or more compliance principles and one or more techniques.

5.5.1 Example 1 Analysis

The first example happened in 2013 and is described in the following excerpt from the Symantec official blog [90]:

> "In April 2013, the administrative assistant to a vice-president at a French-based multinational company received an e-mail referencing an invoice hosted on a popular file sharing service. A few minutes later, the same administrative assistant received a phone call from another vice president within the company, instructing her to examine and process the invoice. The vice president spoke with authority and used perfect French. However, the invoice was a fake and the vice president who called her was an attacker.
>
> The supposed invoice was actually a remote access Trojan (RAT) that was configured to contact a command and control (CC) server located in Ukraine. Using the RAT, the attacker immediately took control of the administrative assistant's infected computer. They logged keystrokes, viewed the desktop, and browsed and ex-filtrated files.

> These tactics, using an e-mail followed up by a phone call using perfect French, are highly unusual and are a sign of aggressive social engineering. In May 2013, Symantec Security Response published details on the first attacks of this type targeting organisations in Europe. Further investigations have revealed additional details of the attack strategy, attacks that are financially motivated and continue to this day."

This example is now mapped to the Social Engineering Attack Framework. It consists of two different phases and also demonstrates how the Social Engineering Attack Framework can handle two different Social Engineering Attacks.

5.5.1.1 First Attack Phase

The important features of the social engineering attack are specified below:

Communication — The Social Engineering Attack is using direct communication with the subclass of unidirectional communication.

Social Engineer — The Social Engineer is an individual.

Target — The Target is an individual. In this instance the target is an administrative assistant to the vice-president at a French-based multinational company.

Medium — The medium is e-mail.

Goal — The goal of the attack is to gain unauthorised access to the organisation.

Compliance Principles — The compliance principles that are used are consistency and authority.

Technique — The technique that is used is phishing.

The next part steps through this example by means of the attack framework.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to gain unauthorised access to the organisation's systems and thus to the organisation's information.

Target identification: The target of the attack is the administrative assistant to the vice-president at a French-based multinational company.

Step 2: Information Gathering

Identify potential sources: Public records of the company and e-mail communication samples from the organisation.

Gather information from sources: Collect and find the public records of the company and collect samples of e-mail communication.

Assess gathered information: Determine the organisational hierarchy and assess the e-mail format of internal organisational e-mail communication.

Step 3: Preparation

Combination and analysis of gathered information: Identify where the target fits into the organisational hierarchy and identify the superiors of the target. Identify the e-mail structure of internal e-mails sent in the organisation and the type of information that should be sent to the target.

Development of an attack vector: Write an e-mail which is similar to other e-mails exchanged within the organisation but also contains the malicious RAT. More specifically, the e-mail's format should be similar to the format used in typical e-mail invoices the administrative assistant receives.

Step 4: Develop Relationship

Establishment of communication: The physical action of sending the e-mail that was developed during the 'development of an attack vector' step is the initial establishment of communication.

Rapport building: The e-mail contents should be similar to a typical e-mail the administrative assistant can expect.

Step 5: Exploit Relationship

Priming the target: The e-mail should be of such a nature that the administrative assistant would not immediately delete or discard the e-mail.

Elicitation: In the 'priming the target' step, the goal is for the target to not delete the e-mail immediately. The elicitation will be deemed successful if the target does not delete the e-mail.

Step 6: Debrief

Maintenance: The e-mail should be worded in such a manner that the target is not perturbed by the e-mail.

Transition: The e-mail should note that there will be some follow-up communication. The target is then prepared for follow-up communication and thus a transition is made to the 'development of an attack vector' step and not to the 'goal satisfaction' step.

5.5.1.2 Second Attack Phase

The important features of the social engineering attack are specified below:

Communication — The Social Engineering Attack is using direct communication with the subclass of bidirectional communication.

Social Engineer — The Social Engineer is an individual.

Target — The Target is an individual. In this instance he is an administrative assistant to the vice-president at a French-based multinational company.

Medium — The medium is the telephone.

Goal — The goal of the attack is to gain unauthorised access to the organisation.

Compliance Principles — The compliance principles that are used are consistency and authority.

Technique — The technique that is used is phishing.

The next part steps through this example by means of the attack framework.

Step 1: Attack Formulation

Nothing here as it is a transition to the 'development of an attack vector' step.

Step 2: Information Gathering

Nothing here as it is a transition to the 'development of an attack vector' step.

Step 3: Preparation

Combination and analysis of gathered information: Nothing here as it is a transition to the 'development of an attack vector' step.

Development of an attack vector: The target already has an e-mail in his inbox containing a malicious invoice, and during phase 1 this e-mail was not deleted. This attack vector is aimed at getting the target to open the malicious invoice so that the social engineer can gain unauthorised access. In this phase one is required to develop a transcript to be followed which will use both authority and consistency principles to get the target to comply with the request to open the malicious invoice.

Step 4: Develop Relationship

Establishment of communication: The physical action of making the phone call of which the transcript has been developed during the 'development of an attack vector' step is the initial establishment of communication.

Rapport building: The telephonic conversation should start off by the attacker introducing himself as the second vice-president of the organisation (This information was gathered from the organisational hierarchy).

Step 5: Exploit Relationship

Priming the target: The target should be aware that the caller requesting him to process the invoice is a person in an authoritative position. It must also be consistent with requests that the target would normally be required to process as well as consistent with the e-mail containing the invoice.

Elicitation: Since the target has been primed to comply with the requests by means of authority and consistency, the social engineer can now request the target to process the malicious invoice.

Step 6: Debrief

Maintenance: The malicious invoice should be similar to one that the target would normally receive. The target should be unaware that he has provided the social engineer with unauthorised access by opening the malicious invoice. Whilst on the phone, the social engineer should be friendly and reassuring towards the target. The target must always feel good about helping the social engineer in order to avoid suspicion.

Transition: The Social Engineer has now obtained his unauthorised access and can proceed to the goal satisfaction state.

Goal Satisfaction: The Social Engineer has obtained his initial goal of obtaining unauthorised access.

5.5.2 Example 2 Analysis

The second example happened in 2009 when fliers appearing to be traffic violations were placed on cars in a parking lot. On these supposed parking violations a website link was included where one could view pictures associated with the so-called violation. The website extracted a Dynamic Link Library (DLL) into the system32 directory on the computer used to access the website. The DLL installs as an internet explorer browser helper object once the system is rebooted. Next a pop-up would appear, informing the user that his computer contains signs of viruses and Antivirus 360 needs to perform a scan. If the user agrees to let the anti-virus application install itself, (it was later found that the anti-virus application was a virus dropper) it in turn installed a virus. The attacker did not continue with the attack, however if he had continued he could have taken full control of the computer since it was already infected with his software [91].

An excerpt of this article reads as follows [91]:

> "I had the opportunity to examine malware whose initial infection vector was a car windshield flier with a website address. The malicious programs were run-of-the-mill; however, the use of fliers was an innovative way of social engineering potential victims into visiting a malicious website.
>
> Several days ago, yellow fliers were placed on the cards in Grand Forks, ND. They stated:
>
> *PARKING VIOLATION This vehicle is in violation of standard parking regulations. To view pictures with information about your parking preferences, go to website-redacted.*
>
> The website showed several photos of cars on parking lots in that specific town. EXIF data in the JPG files show that they were edited using Paint Shop Pro Photo 12 to remove license plate details of the cars and that the photos were taken using a Sony DSC-P32 camera. Installing PictureSearchToolbar.exe led to DNS queries for childhe.com, a domain with a bad reputation according to Symantec, McAfee, etc. Even without the Internet connection, the program installed (extracted) a DLL into C:/WINDOWS/system32."

This example is now demonstrated through the use of the Social Engineering Attack Framework.

5.5.2.1 The Attack Phase

The important features of the social engineering attack are specified below:

Communication — The Social Engineering Attack is using indirect communication through third party mediums.

Social Engineer — The Social Engineer is an individual.

Target — The Target is an individual. In this instance, it is any owner of a car parked in the parking lot.

Medium — The medium is fliers.

Goal — The goal of the attack is to gain unauthorised access to an individuals computer.

Compliance Principles — The compliance principles that are used are social compliance and authority.

Technique — The technique that is used is phishing.

The next part steps through this example by means of the attack framework.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to gain unauthorised access to an unspecified individuals' computer.

Target identification: The target of the attack is any person who owns a car and is parked in the parking lot at the time of spreading the fliers.

Step 2: Information Gathering

Identify potential sources: Public websites with the ability to view parking violations and any institute with authority to reach out a parking violation.

Gather information from sources: Collect sample parking violations which are placed on windshields of cars and sample websites where one can view parking violations.

Assess gathered information: Determine which parking violations are relevant to the specific parking lot, perhaps on location, region etc. The violation, in this case, should specifically conform to the standard parking violations reached out in Grand Forks, ND. Also filter out the website that is consistent with the parking violation.

Step 3: Preparation

Combination and analysis of gathered information: Choose one final parking violation / website pair and finalise the structure of the parking violation, the style and working of the website.

Development of an attack vector: Develop a parking violation consistent to the finalised structure as well as a phishing website which looks similar to the one chosen in the previous step. On the parking violation, ensure that there is a section stating that pictures with information about the parking violation are on a certain website, with a link to the phishing website.

Step 4: Develop Relationship

Establishment of communication: The physical action of putting the created fliers on the cars in the parking lot.

Rapport building: The parking violation placed on the windshield of the cars should be consistent with parking violations handed out in that parking lot under standard conditions. The owner of the car receiving the violation should not doubt whether it is official; it should look legitimate. When the target visits the website, the website should also look legitimate, not raising doubt with the user.

Step 5: Exploit Relationship

Priming the target: The flier should be realistic so that the owner of the car would take it seriously and not just throw it away. While driving home the target should ideally think about the violation and prepare himself to go to the website to view the parking violation, feeling pressured due to social compliance to do the right thing and pay the fine.

Elicitation: Provide a link on the flier which links to the phishing website. Upon clicking on the link, a backdoor is installed on the person's computer, giving the social engineer the opportunity to gain unauthorised access to the computer.

Step 6: Debrief

Maintenance: The flier and website should be created in such a way that the target does not feel threatened. The website should be similar to the real violations website so that the victim is confident that he should take the steps required to pay the violation.

Transition: The social engineer can use the backdoor to gain unauthorised access to the computer and can thus proceed to the 'goal satisfaction' step.

Goal Satisfaction: The Social Engineer has obtained his initial goal of unauthorised access.

5.6 CONCLUSION

The protection of information is extremely important in a modern society and even though the security around information is continuously improving, the one weak point is still the human being who is susceptible to manipulation techniques. This chapter explored social engineering as a domain and social engineering attacks as a process inside this domain.

Kevin Mitnick's social engineering attack cycle [4] is analysed and discussed in detail. The author proposes a social engineering attack framework based on Mitnick's attack cycle. The shortcomings in Mitnick's attack cycle are explored and improvements of these short-comings are reflected in

the proposed attack framework. Each phase in the proposed social engineering attack framework is discussed in detail and two life scenarios are explored as an application of the combination of the attack framework and the previously proposed ontological model.

The author is of the opinion that Mitnick's attack cycle is a good base for social engineering attacks, but lacks significant detail. It is a very broad explanation of an attack and assumes that certain components of the attack are already known, such as the goal of the attack and the target. The attack framework provides specific steps to identify these component and detailed steps for all other aspects of an attack.

This chapter provided an in depth social engineering attack framework as an extension to the proposed ontological model, which was provided in Chapter 2. The framework adds temporal data such as flow and time whereas the ontological model contains all the components of a social engineering attack. The framework and the ontological model can be used to generate social engineering attack scenarios as well as to map historical social engineering attacks to a standardised format. This is important as these scenarios can be used for education and awareness purposes and enables anyone to analyse and compare different social engineering attacks. The following chapter explores the development of such social engineering attack examples.

CHAPTER 6 SOCIAL ENGINEERING ATTACK EXAMPLES

6.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the social engineering attack examples. This chapter is based on a journal article entitled "Social engineering attack examples, templates and scenarios" which has previously been published by the author in the Computers & Security Journal in 2016 [92]. This journal article proposed ten social engineering attack examples which were all derived from examples within existing literature. The purpose of the social engineering attack examples is to have formally constructed attack examples which can be utilised to validate social engineering models, without the need of performing the attack on an individual. In terms of this thesis, the social engineering attack examples are utilised to validate the proposed SEADM, throughout all iterations of the model, without the need to perform the attack and thus limiting the potential harm that may be caused to individuals. The decision was made to validate the SEADM by means of these examples, due to the ethical constraints as provided in Chapter 4. This chapter is provided here as the second contribution chapter as it uses the framework, as proposed in Chapter 5, to develop the social engineering attack examples.

6.2 INTRODUCTION

In order to compare different models, processes and frameworks within social engineering, it is required to have a set of fully detailed social engineering attack examples. Such a set will allow researchers to test their models, processes and frameworks and compare their performances against other models, processes and frameworks.

The author has identified that there are limited practical examples of social engineering in literature. Current literature on social engineering attacks does not depict the full process flow of a social engineering attack and when researchers use these examples, several steps of the attack have to be inferred [87, 91, 90].

This chapter utilises the SEAF, as proposed in Chapter 5, to propose a set of fully detailed social engineering attack examples. Each of the social engineering attack examples is explained by mapping each step of the example to the social engineering attack framework. The examples are extrapolated from existing examples in literature, where the gaps in the literature is completed with elements of what is expected to have occurred.

The remainder of the chapter is constructed as follows. Section 6.3 proposes the social engineering attack examples and maps each example to both the ontological model and the social engineering attack framework. Section 6.4 concludes this chapter.

6.3 EXAMPLES OF A SOCIAL ENGINEERING ATTACK

The author previously proved the usefulness of the social engineering attack framework by mapping well-known social engineering attacks (which have been widely documented in news articles) to the social engineering attack framework. During this research it was found that several pieces of information about the social engineering attack were not included in the documentation and that several steps of the social engineering attack had to be inferred.

The ‘goal identification’ and ‘target identification’ steps are usually not documented. News articles report on an attack after it occurred and the focus is on how the attack affected the specific target. There is also very little information on what steps were followed during the ‘information gathering’ phase. The reader of the news article is to assume that the social engineer performed extensive information gathering on both the goal and the target, which in turn led to a successful social engineering attack. Depending on the type of attack, the ‘preparation’ phase and the ‘develop relationship’ phase normally have information that can be used directly in the social engineering attack framework. The ‘exploit relationship’ phase is not always documented as the specific priming and elicitation techniques are not mentioned specifically. It is normally only mentioned whether the attack was successful or not. The

'debrief' phase is usually also not covered in a report or news article as the 'maintenance' step is a step the social engineer follows to reassure the victim that he/she is not the prey of a social engineering attack. The 'transition' step is something only the social engineer has knowledge of, as the report or news article only reports on the final successful social engineering attack.

The proposed examples attempt to address the problem described above by detailing every phase and associated steps of the social engineering attack framework in such a way that each example will provide repeatable results. The examples are also kept as simple as possible so that they can be expanded upon to create more elaborate examples with exactly the same principal structures. The examples were developed in such a way that other researchers can use them to perform repeatable experiments of social engineering attacks, with repeatable results, without having to physically perform the attack and potentially cause harm to innocent targets [66].

The examples are fairly diverse in order to show and test different social engineering attack scenarios. They are grouped according to the communication type, namely bidirectional communication, unidirectional communication or indirect communication. The classification structure is based on the fact that each example has a specific communication method and that there is almost no overlap of attacks that use the same communication method.

The rest of this section proposes four bidirectional communication examples (Section 6.3.1 to 6.3.4), three unidirectional communication examples (Section 6.3.5 to Section6.3.7) and three indirection communication examples (Section 6.3.8 to Section 6.3.10).

6.3.1 Bidirectional Communication — Example 1

The detailed example of this attack is developed by using elements from the following examples in literature:

- The social engineer (SE) pretends to be someone who works on the management floor and convinces a cleaner of his supposed role. The cleaner grants the social engineer access to the building. This allows the SE to gain physical access to the computerised terminals on the management floor [93, 94].

- The SE pretends to be part of the organisation, dresses in the appropriate attire, and then tailgates into the building behind other employees [89, 37]. This is one of the more difficult attacks to prevent, because people generally feel compelled to hold open the door for other individuals [95, 96].

- The SE can use fake credentials or even just a good story to gain access to an organisation. This can be done by simply printing fake business cards, dressing the part or just carrying the correct security badge [97].

This example illustrates a social engineering attack (SEA) where the attacker attempts to gain physical access to a computerised terminal at the premises of an organisation. The assumption is that when the attacker has once gained access to the computerised terminal, he/she is deemed to have been successful. The attacker is now able to install a backdoor onto the computerised terminal for future and further access from the outside.

The important features of the SEA are specified below:

Communication — The SEA is using bidirectional communication.

Social Engineer — The Social Engineer (SE) is an individual.

Target — The target is an organisation.

Medium — The communication medium is face-to-face.

Goal — The goal of the attack is to gain unauthorised access to a computerised terminal within the organisation.

Compliance Principles — The compliance principles that are used are authority, commitment and consistency.

Techniques — The technique that is used is pretexting.

The following text dissects and maps the example to the Social Engineering Attack Framework (SEAF).

Step 1: Attack Formulation

Goal identification: The goal of the attack is to gain unauthorised access to any computerised terminal within the organisation.

Target identification: The target of the attack is the organisation as a whole. This allows the attacker to target any individual within the organisation who has the capability of allowing the attacker access to the computerised terminal.

Step 2: Information Gathering

Identify potential sources: The information sources include the company website, any individuals who deal directly with the technical support organisation contracted by the target organisation, and information from the technical support organisation gained directly.

Gather information from sources: Gather information from all above mentioned sources that relate directly to how and when technical supported is requested and performed.

Assess gathered information: Determine which technical support company used by the target organisation is most likely to have the authority to gain physical access to the computerised terminal. In addition, determine what time slots can be used to gain physical access to the computerised terminal and whether additional information is required, such as whether the technical support organisation staff must wear corporate uniforms.

Step 3: Preparation

Combination and analysis of gathered information: Determine the best single time slots in which the attacker can attempt to gain physical access to the computerised terminal. This decision will be based on likely time slots during which technical support may be required. The

attacker must also ensure that he is aware of whether corporate uniform is used by the technical support organisation.

Development of an attack vector: Develop an attack plan that contains the exact time the attacker will visit the premises, the precise individual at the premises whom the attacker will ask to gain access to the computerised terminal, and conversation guidelines that should be followed during the attack. The attacker also has the option to perform another SEA in which he can make an appointment for the time slot during which he will attempt to gain unauthorised access to the computerised terminal.

Step 4: Develop Relationship

Establishment of communication: The physical action of engaging the individual within the organisation who can potentially provide the attacker unauthorised access to the computerised terminal.

Rapport building: The attacker is required to develop a friendly relationship with the targeted individual in order for that individual to gain trust in the attacker.

Step 5: Exploit Relationship

Priming the target: The attacker is required to discuss some concerns that he has with the targeted computerised terminal and to prime the targeted individual so that the latter is fully capable and willing to assist with resolving this concern.

Elicitation: The attacker offers to assist in addressing or resolving the concern that the targeted individual experienced with the computerised terminal.

Step 6: Debrief

Maintenance: After the attacker has performed all tasks required on the computerised terminal, he approaches the targeted individual again and assures the latter that all concerns with regard to the computerised terminal have been addressed.

Transition: The attacker was able to successfully gain unauthorised access to the computerised terminal and can thus proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of gaining unauthorised access.

6.3.2 Bidirectional Communication — Example 2

The detailed example of this attack is developed by using elements from the following examples in literature:

- The theory of group conformity is well entrenched in social psychology. The SE uses this theory to his/her advantage by starting a conversation in the group and providing false sensitive information to the group. If most of the other participants in the group are trained by the SE, they also start providing false sensitive information. This will cause any other individual who is part of the conversation to also feel the need to share sensitive information, as he/she will have the ultimate need to belong to the group [98, 99, 100, 101, 102, 103, 104].

- The SE abuses the fact that people feel the need to conform to the group. The SE attempts to convince the target that everyone else has been giving the SE the same information that is now requested from the target [37].

This example illustrates an SEA where the attacker attempts to obtain access to an individual's personal log-on credentials for a specific log-on location. In this case, an attempt is made to gain access to the individual's workstation. The attack will be performed by abusing the psychological principle that an individual has the desire to feel part of a group. Due to commitment and consistency, that individual will feel compelled to conform to what the rest of the group does. In this case, the group of individuals will all reveal their log-on credentials and because the target is the last person in the group to be approached, he/she will feel obliged to also reveal his/her own log-on credentials. The assumption is made that after the attacker has gained the log-on credentials, the SEA is deemed to be successful because these credentials can be used to access the individual's workstation.

The important features of the SEA are specified below:

Communication — The SEA is using bidirectional communication.

Social Engineer — The SE is a group of individuals.

Target — The target is an individual.

Medium — The communication medium is face-to-face.

Goal — The goal of the attack is unauthorised information disclosure from the target to the attacker.

Compliance Principles — The compliance principles that are used are commitment and consistency.

Techniques — The technique that is used is *quid pro quo*.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to get the target to disclose information, which the attacker is not authorised to have.

Target identification: The target of the attack is an individual whose workstation the SE needs to access.

Step 2: Information Gathering

Identify potential sources: The information sources include the places the target visits, any social gatherings the target attends and any interests that the target might have.

Gather information from sources: Gather information from all the above-mentioned sources that relate directly to the specific events the target attends, during which time intervals these events occur and what interests the target has.

Assess gathered information: Determine which of the events the SE is able attend and the length of interaction the SE can have with the target at each of these events. Also, determine how likely individuals will be to interact socially at each of these events and whether the SE will be able to have a conversation with the target at these events.

Step 3: Preparation

Combination and analysis of gathered information: Determine which social events are most likely to present the attacker the possibility to perform an SEA. The events with the highest probability of social interaction and the longest duration with the target should be selected.

Development of an attack vector: Develop an attack plan that contains the chosen event the SE will attend and that states the time interval when the SE will interact with the target. In addition, develop conversational guidelines that will be used during the SEA.

Step 4: Develop Relationship

Establishment of communication: Take the physical action of engaging in conversation with the individual at the chosen event.

Rapport building: The SE, in this case a group of individuals, is required to engage in friendly conversation with the target and make him/her feel part of the group. The SE attempts to build a trust relationship with the targeted individual.

Step 5: Exploit Relationship

Priming the target: After the trust of the target has been gained, the group of individuals is required to steer the conversation onto the topic of password security and how people rarely use complex passwords.

Elicitation: One of the individuals in the group close to the target is required to start off by asking another individual in the group what their log-on credentials are to illustrate that most users use insecure passwords. After the individual has provided his log-on credentials, each of

the other individuals should comply with the request and provide their log-on credentials as well. When all the other individuals in the group have provided their log-on credentials, the target must be requested to provide his log-on credentials. Because of his desire to be part of the group, the target is likely to feel obliged to supply his log-on credentials.

Step 6: Debrief

Maintenance: After the target has provided his log-on credentials, the group should continue with friendly conversation and steer the topic onto some other topic that is of interest to the target. This will have a calming effect on the target and will put him at ease over the fact that he has just released information to which the SE should not have access.

Transition: The attacker was able to successfully persuade the target to disclose unauthorised information and thus the SE can proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of unauthorised information disclosure.

6.3.3 Bidirectional Communication — Example 3

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE pretends to be a network administrator and requests the organisation to provide or reset a user's password on the organisation's system [37].

- The SE gathers information from a third party organisation which can then be used against another organisation [105, 106].

- The SE pretends to be an authoritative figure who is requesting the target to perform a task. Since the target is scared to deny requests from such an authoritative figure, the target may feel compelled to comply with the request [107].

This example illustrates an SEA where the attacker attempts to gain the password of a specific individual's e-mail account where the e-mail account is managed by an organisation. This attack is aimed at the organisation who is in control of the individual's e-mail account and not directly at the individual. Due to this, the individual is considered to be the primary target while the organisation that is targeted is considered a secondary target. The assumption is made that after the attacker has been able to successfully request a password reset for the individual's e-mail account from the organisation, the attacker will be able to gain access to the e-mail account. This is then deemed to be a successful SEA.

The important features of the SEA are specified below:

Communication — The SEA is using bidirectional communication.

Social Engineer — The SE is an individual.

Target — The primary target is an individual. This individual has an e-mail account at a specified organisation, and the latter is considered to be a secondary target.

Medium — The communication medium is a telephone.

Goal — The goal of the attack is to gain unauthorised access to the individual's e-mail account.

Compliance Principles — The compliance principles that are used are authority and scarcity.

Techniques — The technique that is used is pretexting.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to gain unauthorised access to the primary target's e-mail account by requesting a secondary target to have the password for the e-mail account reset.

Target identification: The primary target of the attack is an individual with an e-mail account at the specified organisation. The specified organisation has control over the target's e-mail account and thus an individual at the organisation (which is considered the secondary target) will be persuaded by social engineering to provide access to the primary target's e-mail account. This allows the attacker to target any individual within the organisation who has the capability of allowing the attacker to reset the password of the target's e-mail account.

Step 2: Information Gathering

Identify potential sources: The information sources include the organisation's website, organisational policies and any source that can provide personal information of the primary target.

Gather information from sources: Gather information from all the above-mentioned sources which relate directly to how and when password resets can be requested and what information is required to be provided during the password reset request. This is an example of where the 'information gathering' phase as a whole will be cyclic, because the SE will analyse the information that is required to perform the password reset request and then during the 'assess gathered information' step, it is required to move back to the 'identify potential sources' step to determine from where the additional personal information can be gathered. To keep the attacks as generic and simplistic as possible, this cyclic process is omitted during the description that follows.

Assess gathered information: Determine what process is followed during the password reset request, what information is requested from the individual requesting a password reset, and assess the validity of all gathered personal information of the primary target.

Step 3: Preparation

Combination and analysis of gathered information: Using all the assessed information, determine the best time slots during which a specific staff member of the organisation who has control over the password request process (the secondary target) can be contacted. In addition, it is required to develop a full profile of the primary target's personal information. This profile is

used to ensure that the attacker will be able to answer any questions that the secondary target may direct at the attacker during the password reset request.

Development of an attack vector: Develop an attack plan that contains the exact time that the organisation will be phoned, a full script of the planned telephonic conversation and an organised list of the personal information of the primary target.

Step 4: Develop Relationship

Establishment of communication: The physical action of making the phone call to the organisation, up to the point where the secondary target can assist the attacker with the password reset request.

Rapport building: The attacker is required to develop a friendly relationship with the individual (secondary target) who can assist with the password reset request. The attacker's intention is to get the targeted individual to trust the attacker.

Step 5: Exploit Relationship

Priming the target: The attacker who is impersonating the primary target will explain to the individual (secondary target) that he/she (the attacker) urgently requires to regain access to 'his/her' e-mail account. One example of a way in which a sense of urgency is created is telling the individual how important it is for the attacker to retrieve a specific document from the primary target's e-mail account and that this document is required immediately for some emergency.

Elicitation: The attacker (who is still impersonating the primary target) will request a password reset for the primary target's e-mail account and put forward as the reason for this request that the attacker is using an alternate workstation to access the e-mail account, therefore it does not have the log-on credentials stored.

Step 6: Debrief

Maintenance: After the attacker has successfully requested the password reset, the attacker will profusely thank the individual for the assistance and congratulate him/her on a job well done.

Transition: Since the attacker was able to successfully request a password reset for the primary target's e-mail account, he/she can thus proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of gaining unauthorised access.

6.3.4 Bidirectional Communication — Example 4

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE pretends to be a customer who has in-depth knowledge of the services that an organisation offers. The SE is able to obtain sensitive information from the help-desk staff by bypassing any checks that require authorisation to be granted [93].

- The SE uses the corporate language of the organisation to gain the trust of the other employees [20].

- The SE pretends to be a new employee and requests information from reception [20].

- The SE pretends to be in distress, in a difficult situation or in a life-threatening emergency. The SE calls the targeted department in an organisation and convinces the target that in order to overcome the distress or emergency, his/her request needs to be fulfilled [108].

This example illustrates an SEA where the attacker attempts to obtain sensitive information of an organisation to which only the employees of the organisation have access. The information is not available to members of the public. Once the attacker has been provided with the sensitive information, the SEA is deemed to have been successful.

The important features of the SEA are specified below:

Communication — The SEA is using bidirectional communication.

Social Engineer — The SE is an individual.

Target — The target is an organisation.

Medium — The communication medium is e-mail.

Goal — The goal of the attack is unauthorised information disclosure from the target to the attacker.

Compliance Principles — The compliance principles that are used are friendship and liking.

Techniques — The technique that is used is pretexting.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to get an employee of the organisation to disclose to the attacker information that the attacker is not authorised to have.

Target identification: The target of the attack is the organisation as a whole. This allows the attacker to target any individual within the organisation who has the sought-after capability of providing the attacker with the sensitive information.

Step 2: Information Gathering

Identify potential sources: The information sources include the organisation's website, any individuals in the organisation who have access to the information, and any organisational policies and procedures.

Gather information from sources: Gather information from all above-mentioned sources that relate directly to the access level of each employee and his/her status in the organisation.

Assess gathered information: Determine which of the employees have access to the sensitive information that the attacker is trying to obtain. Also, assess all the gathered information about each employee and perform information gathering on each of the employees individually. This cyclic process is excluded from the example and it is assumed that for the next phase all personal information about each employee has been gathered and assessed.

Step 3: Preparation

Combination and analysis of gathered information: Determine the level of susceptibility of each employee, how much access to information each employee has and what type of personal information the attacker was able to gather and assess about him/her. Also, develop an information profile on each employee to determine which employee would be the best target from whom to request the sensitive information.

Development of an attack vector: Develop an attack vector that contains the chosen employee whom the attacker will be targeting, the full personal profile of this employee and what level of access this employee has. In addition, develop the planned e-mail communication with the employee to fit the specific personal profile of the employee.

Step 4: Develop Relationship

Establishment of communication: The very first e-mail communication that the attacker has with the targeted employee of the organisation. This e-mail establishes the basis for all future communication between the attacker and employee.

Rapport building: This step will be a continuous process of back and forth e-mail communication between the attacker and the employee. Several e-mails will be transferred in a bidirectional manner between the attacker and the employee in order to gain the trust of the employee. An example of trust building is where the attacker appears to be interested in the hobbies and interests of the targeted employee. The similarity between the attacker and the targeted employee's preferences is used to build trust.

Step 5: Exploit Relationship

Priming the target: The exploitation of the relationship will occur within a single e-mail communication to the targeted employee. In the priming and elicitation e-mail, the attacker will inform the employee of a scenario in which the attacker requires access to the sensitive information. An example of this could be that the attacker is requesting sensitive information about the company policies because the attacker, as part of the pretext, will be attending an interview at the targeted employee's organisation.

Elicitation: The attacker will request the assistance of the targeted employee to retrieve the sensitive information and due to the friendship and liking and the trust relationship that have been established, the targeted employee will feel obliged to comply with the request.

Step 6: Debrief

Maintenance: It is important that the attacker does not abruptly end the communication between himself and the targeted employee as this may cause suspicion and the organisation may be alerted to a breach of information. The attacker is required to continue the e-mail communication until such time as the request that was made is likely to have been forgotten by the targeted employee and the topic of communication has moved on away from the information request. The e-mail communication should thus continue until the sensitive information has been utilised by the attacker and is no longer of use.

Transition: The attacker was able to successfully gain unauthorised information disclosure from the targeted employee and can thus proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of unauthorised information disclosure.

6.3.5 Unidirectional Communication — Example 1

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE deploys a fake website that sells tickets for a sporting event. The SE also sends out phishing e-mails to inform people that they can buy discounted tickets [93].

- The SE sends out phishing e-mails that falsely originate from the e-mail addresses of known contacts. Due to the targeted nature of the phishing attempts, the success ratio increases significantly [109].

- The SE sends out an e-mail that directs the target to navigate to a fraudulent website, which in turn collects credentials such as identity document numbers and bank account numbers from the target [22].

This example illustrates an SEA where the attacker attempts to obtain financial gain by sending out e-mails that request a group of individuals to make a small deposit into a bank account owned by the attacker. The '419 scams', which are very popular social engineering attacks, are examples of this type of attack. Once the attacker has received the small deposit from the targeted individual, the SEA is deemed to have been successful.

The important features of the SEA are specified below:

Communication — The SEA is using unidirectional communication.

Social Engineer — The SE is an individual.

Target — The target is a group of individuals.

Medium — The communication medium is e-mail.

Goal — The goal of the attack is financial gain, as the targets are requested to make a deposit into a bank account owned by the attacker.

Compliance Principles — The compliance principle that is used is scarcity.

Techniques — The technique that is used is phishing.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to get an individual to deposit money into a bank account owned by the attacker and thus to provide financial gain to the attacker.

Target identification: The target of the attack is any individual of which the attacker has an e-mail address.

Step 2: Information Gathering

Identify potential sources: The information sources include any publicly available e-mail lists, websites selling e-mail lists and any other locations that are used to store e-mail addresses.

Gather information from sources: Gather from all the above-mentioned sources information that relates directly to the individuals' personal information and e-mail addresses.

Assess gathered information: Determine whether each e-mail list that has been gathered contains all information about each individual and whether each individual has an associated e-mail address.

Step 3: Preparation

Combination and analysis of gathered information: Combine all the lists obtained into a single list that contains the personal details of each individual and his/her associated e-mail address. After the lists have been combined, prune all duplicates from the list to create a single list with only unique e-mail addresses.

Development of an attack vector: Develop an attack plan that details all the information that should be contained in each e-mail, what personal information to use in each e-mail and exactly how each section of the e-mail should be worded. It is also important to determine the duration

of the attack, because the attacker will have to close the bank account after a specified amount of time to ensure that individuals are not able to reverse any funds transferred.

Step 4: Develop Relationship

Establishment of communication: This involves the physical action of sending out an e-mail to each of the e-mail addresses on the list.

Rapport building: Rapport building in an e-mail usually occurs already in the subject line and in the first few paragraphs of the e-mail. The reason behind this is that individuals scan only the subject line and the first few paragraphs of an e-mail, and trust should be built so that the target is enticed to read the entire e-mail.

Step 5: Exploit Relationship

Priming the target: In this attack, priming is done by using the scarcity principle. Priming usually occurs in the paragraphs following the 'rapport building' step. In these paragraphs, the target is informed that he/she is a specially selected individual and that there is only a limited time frame within which to claim the reward offered to him/her in this e-mail.

Elicitation: In the next paragraph, the attacker requests the individual to make a smaller deposit than the reward offered, in order to be eligible to claim the full reward.

Step 6: Debrief

Maintenance: The e-mail is ended off by thanking the target so as to make him/her feel at ease about making the payment and being selected for the specific reward.

Transition: If the attacker is successful in his/her request that the target makes a payment into the attacker's bank account, the attacker can proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of financial gain.

6.3.6 Unidirectional Communication — Example 2

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE utilises a pop-up-window attack that is deployed on the user's workstation. When the user logs on to the specific service for which the SE requires the user's log-on credentials, a pop-up window can appear that requires the user to repeat his/her log-on credentials [110].

- The SE also uses a pop-up-window attack while the user is logged into a system. The SE lets the workstation show a pop-up window that informs the user that the specific application has had a problem and that the user is required to re-authenticate. This re-authentication dialogue box then captures the user's log-on credentials and provides them to the SE [49].

- The SE sends the target a message by using a mobile device. The message indicates that the user has to update the application that is used to access the system or the product to which the user has access. This can convince the user to visit the link and during the update process, the user is asked to provide his/her log-on credentials [111].

This example illustrates an SEA where the attacker attempts to obtain log-on credentials from a group of individuals who are all using a certain system or product provided by an organisation. It is assumed that individuals are required to log-on to this system or product using log-on credentials unique to each individual. Individuals who are using the system are not allowed to share their log-on credentials and thus the goal of this attack is unauthorised information disclosure. The SE can have a further goal to obtain unauthorised access to the system or product, but that is seen as a separate goal. Once the attacker has obtained the log-on credentials from the individual, the SEA is deemed to be successful.

The important features of the SEA are specified below:

Communication — The SEA is using unidirectional communication.

Social Engineer — The SE is an individual.

Target — The target is a group of individuals.

Medium — The communication medium is a SMS.

Goal — The goal of the attack is unauthorised information disclosure from the target to the attacker.

Compliance Principles — The compliance principles that are used are scarcity, commitment and consistency.

Techniques — The technique that is used is phishing.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to get an individual to provide to the attacker information that the attacker is not authorised to have.

Target identification: The target of the attack is all individuals in the group who are using the system provided by an organisation.

Step 2: Information Gathering

Identify potential sources: The information sources include any information about the system, the organisation's website and any lists that contain details of the users of the system.

Gather information from sources: Gather from all the above-mentioned sources information that relates directly to the individuals' personal information, cellphone numbers and any information regarding the product and the appearance of the log-on screen for the product.

Assess gathered information: Determine whether each identified user has an associated cellphone number and that the cellphone number is valid. Also, assess if enough information has been gathered to correctly duplicate the log-on screen for the specific system.

Step 3: Preparation

Combination and analysis of gathered information: Develop a single list that contains the names of all users of the system and their associated cellphone numbers. In addition, develop a mock-up of how the log-on screen should look, so that this can be replicated to ensure that the screen is familiar to the targets during the attack.

Development of an attack vector: Develop an attack plan that details all the information that should be contained in each SMS, what personal information to use in each SMS and exactly how each section of the SMS should be worded. For this example, the attackers are required to develop a log-on screen that looks similar to the original screen and that is able to capture the log-on credentials when individuals attempt to log-on.

Step 4: Develop Relationship

Establishment of communication: This is done by the physical action of sending out all the SMSs to each of the cellphone numbers on the list.

Rapport building: Rapport building in an SMS usually occurs in the very first sentence of the SMS. The reasoning behind this is that SMSs are limited to 160 characters and thus you are required to keep the content brief. The first sentence of the SMS should build trust in the individual and entice him/her to read the rest of the SMS. In this example, the SMS would mention that it is an automated SMS from the organisation providing the system.

Step 5: Exploit Relationship

Priming the target: The second sentence of the SMS is used to both prime the target and to elicit action. The attacker will prime the target by using the scarcity principle, and by saying that a free update for the system will be available for a limited period only.

Elicitation: The sentence continues by providing a shortened hyperlink in the SMS on which the individual will be requested to click to obtain the free update to the system. The first screen that the individual would see after clicking on the link would be a log-on screen similar to what

he/she is used to. Using the commitment and consistency principles, the user will trust the familiar-looking site and enter his/her log-on credentials.

Step 6: Debrief

Maintenance: In this example, maintaining rapport is actually performed on the log-on screen and not in the SMS itself. After the user has logged on to the fraudulent system, a message appears thanking the individual for updating to the latest version and the individual is then redirected to the original system.

Transition: The attacker was able to successfully gain unauthorised information from the target and can thus proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of unauthorised information disclosure.

6.3.7 Unidirectional Communication — Example 3

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE performs a pretext using postal letters. The SE pretends to be various officials, internal employees, employees of trading partners, customers, utility companies or financial institutions and the SE solicits confidential information by using a wide range of persuasive techniques [11].

- The SE has the capability of spoofing the sender ID on popular mobile messaging applications [112]. This capability can further be used to perform an SEA and to send messages to other users whilst impersonating friends of these users [113].

- Typical SE attacks, specifically phishing, used to occur via postal mail. The term '419 scams' refers to section 419 of the Nigerian Criminal Code, which outlaws this type of scam. During the 1970s, postal mail was mostly used in these scams and during the 1980s, the medium

of communication changed to faxes. Both are examples of forms used by the SE to initiate unidirectional communication [114].

This example illustrates an SEA in which the attacker attempts to obtain financial gain by sending out paper mail. This letter requests a group of individuals to make a small deposit into a bank account owned by the attacker. In this example, the attacker develops a phishing letter that masks the attacker as a charity organisation requesting donations. Once the attacker has received the small deposit from the targeted individual, the SEA is deemed to be successful.

The important features of the SEA are specified below:

Communication — The SEA is using unidirectional communication.

Social Engineer — The SE is an individual.

Target — The target is a group of individuals.

Medium — The communication medium is paper mail.

Goal — The goal of the attack is financial gain because the targets are requested to make a deposit into a bank account owned by the attacker.

Compliance Principles — The compliance principle that is used is scarcity.

Techniques — The technique that is used is phishing.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to get an individual to make a deposit into a bank account owned by the attacker and thus allowing the attacker to achieve financial gain.

Target identification: The target of the attack is any individual for whom the attacker has a postal address.

Step 2: Information Gathering

Identify potential sources: The information sources include any publicly available telephone records and address lists.

Gather information from sources: Gather from all the above-mentioned sources information that relates directly to the individuals' personal information and postal address.

Assess gathered information: Determine whether each address list that has been obtained contains all information about each individual and whether each individual has an associated postal address.

Step 3: Preparation

Combination and analysis of gathered information: Combine all the lists obtained into a single list that contains the personal details of each individual and his/her associated postal address. After the lists have been combined, prune all duplicates from the list to create a single list with only unique postal addresses.

Development of an attack vector: Develop an attack plan that details all the information that should be contained in each letter, what personal information to use in each letter and exactly how each section of the letter should be worded. It is also important to determine the duration of the attack, as the attacker will have to close the bank account after a specified amount of time to ensure that individuals are not able to reverse any funds transferred.

Step 4: Develop Relationship

Establishment of communication: This is done by the physical action of sending out letters to each of the postal addresses on the list.

Rapport building: Building rapport in postal mail is very similar to building rapport in an e-mail and it should occur in the first few paragraphs of the letter. In this example, the first few paragraphs should introduce the charity requesting the donation and what the charity has done so far with previous donations received. This information is used to build trust in the individual and to ensure that the individual will support the charity and want to read the rest of the letter.

Step 5: Exploit Relationship

Priming the target: The individual is primed by providing him/her with a list of the current donations that have been received by the charity, what the charity needs to purchase and specifically why these donations will be needed. The received donations section will assure the individual that there are other people donating and that it is socially acceptable to donate to the charity. The additional work the charity can perform and why the donations are requested are included to provoke an emotional response from the individual so that he/she can relate to the charity.

Elicitation: Using an empathetic tone of writing, the attacker requests the individual to make a small donation to the specified charity. It is very important to provide several options on how the individual can donate to the charity and the procedure to perform the donation should be as simple as possible.

Step 6: Debrief

Maintenance: The letter is finalised by thanking the individual for his potential generosity and to assure the individual that any donation that is made will be spent wisely.

Transition: If the attacker succeeds in persuading the target to make a payment into the attacker's bank account, the attacker can proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE is satisfied as he/she attained the initial goal of financial gain.

6.3.8 Indirect Communication — Example 1

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE scatters USB drives in the parking lot, smoking areas and other areas that employees frequent. The employees plug in the USB drives the minute they get to their workstations [115].
- The SE attempts to gain unauthorised access to a workstation in an organisation by using a storage medium or device [116, 117].
- Spreading malware through means of storage media or storage devices is nothing new; this practice can be traced back to the use of floppy drives [22].

This example illustrates an SEA in which the attacker attempts to gain unauthorised access to a workstation within an organisation by using a storage device. Once the target has plugged the storage device (in this example a USB flash drive) into the targeted workstation, the SEA is deemed to be successful. This is because the attacker is now able to install a backdoor onto the workstation via the storage device. The SE can then proceed to use this workstation as a pivot point for any further attacks on the organisation.

The important features of the SEA are specified below:

Communication — The SEA is using indirect communication.

Social Engineer — The SE is an individual.

Target — The target is an organisation.

Medium — The communication medium is a storage device. In this example, the storage device to be used is a USB flash drive.

Goal — The goal of the attack is to gain unauthorised access to a workstation within the organisation.

Compliance Principles — The compliance principle that is used is social validation.

Techniques — The technique that is used is baiting.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to gain unauthorised access to any workstation within the organisation.

Target identification: The target of the attack is the organisation as a whole. This allows the attacker to target any individual within the organisation who has a workstation or who has access to a workstation.

Step 2: Information Gathering

Identify potential sources: The information sources include physical scouting of the premises, monitoring of the movement of employees, and any schedules or appointments posted on the organisation's website.

Gather information from sources: Gather from all the above-mentioned sources information that relates directly to how and when employees are entering and leaving the office building and specifically which entrances are being used.

Assess gathered information: Determine which of the entrances are the most viable target, based on the time intervals when individuals enter and exit the organisation at these entrances. Also, determine the possible ways the attacker can approach these entrances without looking suspicious or showing suspicious behaviour.

Step 3: Preparation

Combination and analysis of gathered information: Determine the best time slots during which the attacker can attempt to deploy the storage medium at the entrance without having to perform any suspicious behaviour. It is important to choose a time slot when most individuals are entering the building, because it is always possible that an individual exiting the building may also pick up the storage medium.

Development of an attack vector: Develop an attack plan that contains the exact time that the attacker will visit the premises, which entrance the storage medium will be deployed at, how the storage medium will be marked to prompt the individual to return it to its owner and what data will be deployed onto the storage medium. The storage medium should contain a Trojan (malware) that will attempt to connect to the attacker's network infrastructure.

Step 4: Develop Relationship

Establishment of communication: Communication is established via the physical action of deploying the storage medium at an entrance and it lasts up to the time when an individual picks up the storage medium.

Rapport building: In this example, rapport is developed by ensuring that the storage medium looks similar to those that are typically used by the organisation and that are branded with the organisation's logo.

Step 5: Exploit Relationship

Priming the target: Attach a label to the storage medium that states that the information on the storage medium is very valuable and that, if lost, it should be returned to the owner. The label or sticker to convey this message is normally only a sticker saying 'Important' or 'Confidential'. The target is required to plug the storage medium into a workstation in order to determine the owner.

Elicitation: The 'elicitation' step is almost implicit in this example. Most people will attempt to return lost valuables or they could just be curious to find out what information is stored on the storage medium. Both of these situations will lead to a successful 'elicitation' step.

Step 6: Debrief

Maintenance: Once the storage medium has been connected to a workstation, the Trojan will automatically execute in a hidden fashion. In order to avoid suspicion, it is good practice by the attacker to include either contact details to return the storage medium or an encrypted document to indicate the importance of the information.

Transition: Once the attacker was able to successfully gain unauthorised access to the workstation of the individual, he/she can proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his/her initial goal of gaining unauthorised access.

6.3.9 Indirect Communication — Example 2

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE studies the available attributes on public profiles within specific social networks and determines how they may be exploited. Context-aware e-mail spam is then generated and sent to users of the network [118]. This same attack can be repeated by posting the context-aware spam within the social networks of the users.

- Users of social networking websites exhibit a high degree of trust in both friend requests and messages from other users. This research also covers reverse social engineering attacks where the victim initiates the conversation with the attacker. [119].

- The SE creates a fake profile that propagates click-bait posts that all use shortened forms of the URL. This lets unsuspecting victims click on the links, which can lead them to websites containing malware [50].

This example illustrates an SEA where the attacker attempts to obtain log-on credentials from a group of individuals who are all using a certain social media website. It is assumed that individuals are

required to log-on to this website using log-on credentials unique to each individual. Individuals who use the particular social media website are not allowed to share their log-on credentials and thus the goal of this attack is unauthorised information disclosure. The SE may have a further goal, namely to obtain unauthorised access to the individual's social media account, but that is seen as a separate goal. Once the attacker has obtained the log-on credentials from the individual, the SEA is deemed to be successful.

The important features of the SEA are specified below:

Communication — The SEA is using indirect communication.

Social Engineer — The SE is an individual.

Target — The target is a group of individuals.

Medium — The communication medium is via a website. In this specific case, it is a social media website.

Goal — The goal of the attack is unauthorised information disclosure from the target to the attacker.

Compliance Principles — The compliance principles that are used are social validation and friendship and liking.

Techniques — The technique that is used is baiting.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to get an individual to provide to the attacker information that the attacker is not authorised to have.

Target identification: The target of the attack is all individuals in the group who are using the specific social media website.

Step 2: Information Gathering

Identify potential sources: The information sources include any information about the social media website, the users of the social media website and the policies of the social media website.

Gather information from sources: Gather from all the above-mentioned sources information that relates directly to the individuals' personal information and any information regarding the log-on page of the social media website.

Assess gathered information: Determine whether all the required information to determine the likes and dislikes of each individual have been gathered. Also, assess if enough information has been gathered to correctly duplicate the log-on screen for the social media website.

Step 3: Preparation

Combination and analysis of gathered information: Develop a combined personality profile based on all the information gathered from the individuals and determine what type of social media posts will be of interest to these individuals. Also, develop a mock-up of how the log-on screen should look, so that the replicated log-on screen looks familiar to the individuals when they are required to enter their log-on credentials during the attack.

Development of an attack vector: Develop an attack plan that details the formulation of a post on which most of the individuals will click, based on their personality profile. In this example, the attacker is also required to develop a log-on screen that is similar to the original, and that is able to capture the log-on credentials when individuals attempt to log-on. Once an individual has fallen prey to the attack, each target that has been compromised by the malicious post will be forced — unbeknown to the target — to automatically replicate the attacker's social media post to that of the target's friends.

Step 4: Develop Relationship

Establishment of communication: This involves the physical action of posting the first social media post on the social media website.

Rapport building: Posts on social media websites are usually very short and often consist of one or two sentences only. The 'rapport building' step is mostly performed as a continuous process because individuals trust people with whom they have been friends on social media for a long period. In this example, the first post by the attacker should be enticing enough for any of the targets to click on it without having gained a lot of trust in the attacker. Once a single individual has fallen prey to the attack, he/she will automatically, due to the malicious post, propagate the post to his/her social media friends, seeing that a trust relationship already exists between friends.

Step 5: Exploit Relationship

Priming the target: On social media websites, the target is almost already primed to be reading and clicking on posts. Individuals usually tend to read social media to find interesting activities that their friends are participating in or have posted. In the post that the attacker provides, the image that accompanies the post and the short content description represent both the 'priming the target' and the 'elicitation' steps.

Elicitation: The post that is made available by the attacker contains both an image and a short description based on the personality profile of the group of individuals who are being targeted. These individuals should be interested in the subject matter that is posted and thus they would hardly hesitate to click on the post and read more about it. Once the individual has clicked on the post to read it, it will ask the individual for his/her log-on credentials for the particular social media website as if he/she has been logged out. The individual is then prompted to log back in to the social media website, after which the post is propagated to all of the target's social media friends.

Step 6: Debrief

Maintenance: In this example, the maintenance of rapport actually occurs on the log-on screen and not in the post made on social media. After the targeted individual has logged on to the

fraudulent log-on screen, the information that was discussed in the fraudulent post should be provided, after which the individual is navigated back to the real social media website. This allows the targeted individual to think that he/she gained access to the post that he/she wanted to read and the target remains unaware that all his/her social media friends have also been posted the fraudulent post.

Transition: The attacker was able to successfully gain unauthorised information from the target and can thus proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his/her initial goal of unauthorised information disclosure.

6.3.10 Indirect Communication — Example 3

The detailed example of this attack is developed by using elements from the following examples in literature:

- The SE creates fake traffic violation notices and places them onto cars at a parking lot. The owner of the car returns to his/her car, finds the notice and later navigates to the URL provided on the traffic violation notice. In this way the owner of the car is tricked to visit a malicious website. This example is directly expanded from the example quoted by [91].

- The SE prints posters that contain a QR code. The poster is then placed close to a popular restaurant and mentions that scanning this QR code with your phone provides you access to a voucher for the restaurant. Upon scanning the code, the QR code directs the target to a malicious website or requests a signup to harvest usernames and passwords [120].

- The SE creates a URL that points to malicious malware on a cloud-based system [121]. This URL is printed on a pamphlet and provided to job seekers who seek employment. The pamphlet advertises a job opportunity and provides a URL to a website where additional information can be found, or where the job seeker must apply.

This example illustrates an SEA in which the attacker attempts to gain unauthorised access to any individual's computer. In the current example, fliers appearing to be fines for traffic violations are placed on different individuals' cars in a parking lot. On these notices of supposed parking violations a website Uniform Resource Locator (URL) is provided where one could view pictures associated with the so-called violation. When the individual visits the website, a backdoor Trojan is installed onto the individual's workstation. Once the individual has accessed the malicious website, the attacker successfully installs the backdoor Trojan and that SEA is deemed to be successful.

This example is now demonstrated through the use of the SEAF.

The important features of the SEA are specified below:

Communication — The SEA is using indirect communication through third-party media.

Social Engineer — The SE is an individual.

Target — The target is an individual. In this instance, it is any owner of a car parked in the parking lot.

Medium — The communication medium is a flier.

Goal — The goal of the attack is to gain unauthorised access to an individual's computer.

Compliance Principles — The compliance principles that are used are social compliance and authority.

Techniques — The technique that is used is phishing.

The following text dissects and maps the example to the SEAF.

Step 1: Attack Formulation

Goal identification: The goal of the attack is to gain unauthorised access to an unspecified individual's computer.

Target identification: The target of the attack is any person who owns a car and is parked in the parking lot at the time when the fliers are spread.

Step 2: Information Gathering

Identify potential sources: Public websites that provide the feature to view parking violation details and any institute with the authority to issue a parking violation.

Gather information from sources: Collect sample parking violation notices that are placed on windshields of cars and on sample websites where one can view parking violation information.

Assess gathered information: Determine which parking violations are relevant to the specific parking lot, perhaps on location, region, etc. In this case, the violation should specifically conform to the standard parking violations that occur in the target region. Also filter out the website that is consistent with the parking violation.

Step 3: Preparation

Combination and analysis of gathered information: Choose one parking violation and website pair and finalise the structure of the parking violation notice, the style and working of the website.

Development of an attack vector: Develop a parking violation notice consistent with the finalised structure as well as a phishing website that looks similar to the one chosen in the previous step. On the parking violation notice, ensure that there is a section stating that photos with information about the parking violation are on a certain website, with the URL of the phishing website.

Step 4: Develop Relationship

Establishment of communication: This is done via the physical action of placing the created fliers on the cars in the parking lot.

Rapport building: The parking violation notices placed on the windshields of the cars should be consistent with parking violation notices handed out in that parking lot under standard conditions. The owner of the car receiving the violation notice should not doubt whether it is official; it should look legitimate. When the target visits the website, the website should also appear to be legitimate and may not raise doubt with the user.

Step 5: Exploit Relationship

Priming the target: The flier should be realistic so that the owner of the car will take it seriously and not simply throw it away. While driving home, the target should ideally think about the violation and prepare himself to go to the website to view the parking violation, feeling pressured due to social compliance to do the right thing and resolve the violation.

Elicitation: The attacker provides a URL on the flier of the phishing website to allow the target to take action. Upon typing in the URL, a backdoor is installed on the target's computer, giving the SE the opportunity to gain unauthorised access to his/her computer.

Step 6: Debrief

Maintenance: The flier and website should be created in such a way that the target does not feel threatened. The website should be similar to the real violations website so that the victim is confident that he/she is performing the correct procedure to resolve the violation.

Transition: The SE can use the backdoor to gain unauthorised access to the computer and can thus proceed to the 'goal satisfaction' step.

Goal satisfaction: The SE has attained his initial goal of gaining unauthorised access.

6.4 CONCLUSION

This chapter revisited both the ontological model and the social engineering attack framework in order to further expand the social engineering domain. The author found that reports and news articles on SE do not provide all the information on social engineering attacks. There is usually no information available on either the 'attack formulation' phase or the 'information gathering' phase. There is also very little information on the 'exploit relationship' phase, because reports or news articles tend to mention only the technique that was used and that it was successful. In order to perform comparative studies of social engineering models, processes and frameworks, it is essential to have a set of fully detailed social engineering attack examples.

This chapter proposed ten examples of social engineering attacks. These examples are designed to be diverse and unique so that there is little overlap between each of them. The examples are also categorised based on the type of communication that is utilised. The author proposes four examples in which bidirectional communication is used, three for unidirectional communication and three for indirect communication.

These proposed social engineering attack examples form a basis for the testing and validation of the SEADM which is proposed in the following chapters. Additionally, these attack examples can be used as a resource by other researchers to expand on, used for comparative measures, creating additional examples or evaluating models for completeness. The following chapter revisits the SEADM, provided in Chapter 3, by addressing the shortcomings of the first iteration.

CHAPTER 7 SOCIAL ENGINEERING ATTACK DETECTION MODEL

7.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the second iteration of the Social Engineering Attack Detection Model (SEADM). Each iteration of the SEADM was presented in a formal publication. These publications focused on addressing the shortcomings that were identified as the research regarding the SEADM matured throughout the research process. This chapter is based on a conference paper entitled "Social engineering attack detection model: SEADMv2" which has previously been published by the author at the Cyberworlds conference in 2015 [122]. This conference paper revisited the first iteration of the SEADM, which was provided in Chapter 3, and attempts to address the shortcomings thereof and expand on the initial model. The SEADM is the main contribution of this thesis, however, a significant amount of shortcomings within the field of social engineering was identified through the research process. All of the previous chapters attempted to address these shortcomings by means of research contributions. The SEADM, in itself, also went through several iterations and this chapter focuses on the second iteration only.

7.2 INTRODUCTION

The first implementation of the SEADM, as provided in Chapter 2, proposed a social engineering attack detect model that was specifically aimed at a call centre environment [44]. Although the model worked well, it catered only for social engineering attacks that utilised bidirectional communication. After further research, the author realised that social engineering attacks can be divided into three categories based on the type of communication, namely bidirectional communication, unidirectional

communication and indirect communication. Hence the author realised the need for a model to cater for all three categories and thus this chapter is dedicated to proposing a revised version of the SEADM.

The problem at hand is to successfully detect social engineering attacks while working in a stressful environment where decisions must be made instantaneously. It is for this reason that a practical model that can be easily implemented and used by all levels of employees is necessary and proposed in this thesis. This model should be used in combination with training on various social engineering techniques, the psychological vulnerabilities they may elicit, and institutional policies and procedures.

The remainder of the chapter is constructed as follows. Section 7.3 provides a background on the previous social engineering model and proposes the revised version of the SEADM. Section 7.4 maps social engineering attack examples to demonstrate the use of the SEADM. Section 7.5 concludes this chapter.

7.3 REVISED SOCIAL ENGINEERING ATTACK DETECTION MODEL

The previous SEADM, SEADMv1, was designed to cater specifically for social engineering attacks in a call centre environment [44, 45]. This research was the first attempt to develop a detection model for social engineering attacks, and at the time of publishing this article there was still only limited research available in this field. Most of the research in this domain still centres around the training of users [59, 56, 123]. The steps in the revised SEADM, SEADMv2, have been generalised to cater for all three communication categories, whereas the previous model was only able to deal with bidirectional communication. This has led to the author developing a revised SEADM which is able to thwart social engineering attacks from all three categories of social engineering.

The author proposes a revised SEADM as depicted in Figure 7.1 depicts the revised SEADM which is able to detect attacks from all three communication categories. This model makes use of a decision tree and breaks down the process into more manageable components to aid decision making.

The model also depicts the flow of action and how any type of request should be handled by a ‘receiver’. Throughout this discussion this term is understood as the person dealing with the request, while the

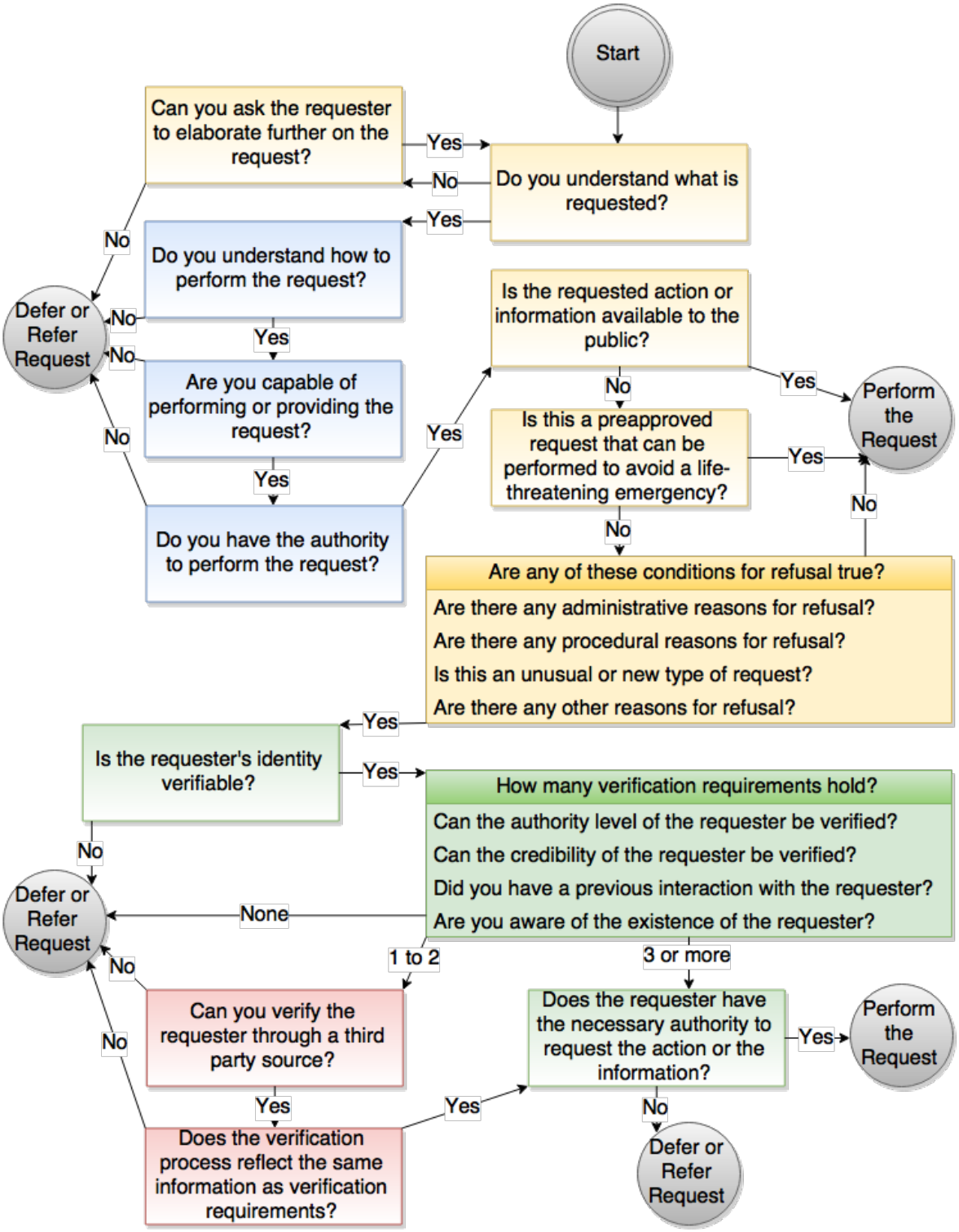

Figure 7.1: Social Engineering Attack Detection Model

term ‘requester’ is defined as the person or object who requests the specific action or information from the receiver. The model should be used as a guideline to aid in decision making and it is an improvement on the initial SEADM due to its ability to cater for both typical requests and inherent requests. This generalisation allows the revised SEADM to cater for the both the unidirectional communication and indirect communication categories of social engineering.

An example of a typical request is where the requester, in this case a person, requests the receiver to perform a task/favour for him/her. This request can range from the requester requesting information about an organisation to the requester requesting that the receiver performs a password reset for an individual’s Internet banking logon.

An example of an inherent request is where the receiver receives a request, in this case an object that contains either a request or a process which needs to be completed by the receiver. This type of request can range from a parking ticket detailing how to pay the ticket on the pamphlet to a receiver finding a storage medium device and wanting to return the device to its rightful owner. In the case of the parking ticket, the receiver is inherently requested to pay the ticket using the information on the pamphlet. In the case of the storage device the situation is a little bit more complicated. The receiver, upon finding the device, is inherently requested to return it to its rightful owner.

The model provides for four different types of states — the request, receiver, requester and third party — which provide a brief idea of what can be expected to be performed in each state. The request states, indicated in yellow, directly deals with information about the request itself. The receiver states, indicated in blue, directly deals with the person handling the request and whether this person (the receiver) understands and is allowed to perform the request. The requester states, indicated in green, directly deals with the requester and whether any information about the requester can be verified. The third party states, indicated in red, directly depict the involvement of a third party in the model and whether the information about the requester can be externally verified.

This thesis addresses each of these states individually as shown in Figure 7.1, before the full model is discussed with examples.

7.3.1 Do you understand what is requested?

The first step of the model tests whether the receiver is able to understand the request. One must fully comprehend the entire request before one can accurately determine whether it involves a social engineering attack or not. In the case where the receiver does not comprehend the request in full or where additional information is required to comprehend the request, the 'no' option is selected so that more information about the request can be asked. The 'no' option can be selected repeatedly if the additional information provided is not yet enough for the receiver to comprehend the request fully. When the receiver fully comprehends what the request is, the 'yes' option is selected.

7.3.2 Can you ask the requester to elaborate further on the request?

This step is provided to allow the receiver to request additional information about the request from the requester. The receiver now attempts to ask for more information about the request to better understand the request. The step can also be repeated if additional information is required; however, when no additional information about the request can be obtained, the 'no' option is selected. The 'no' option leads the receiver to either defer or refer the request. Every time more information is obtained, the 'yes' option is selected and a test is once again performed to determine whether the receiver understands what is requested.

7.3.3 Do you understand how to perform the request?

It is very important to determine whether the receiver is knowledgeable enough to perform the request in full. This step differs from the first step, because the receiver has to measure whether he/she has the required knowledgeable to perform the request and not whether the request itself is understood. The receiver performs this measurement based on the own capability to understand the procedure required to perform the request. If the receiver has the necessary knowledge to perform the request, the 'yes' option is chosen and the receiver proceeds to the following step. In the case where the receiver does not understand how to perform the request, the 'no' option is chosen.

7.3.4 Are you capable of performing or providing the request?

Once the receiver has verified that he/she understands how to perform the request, it needs to be determined whether the receiver is capable of performing the request. This step measures whether the receiver has the means or capability to perform the request, in other words whether he/she is able to determine which option to choose. If the 'no' option is taken, the request is deferred or referred to another individual. Otherwise, the 'yes' option is selected and the next step involves the receiver's authority to perform the request.

7.3.5 Do you have the authority to perform the request?

The receiver needs to measure whether he has the required authority to perform the request. Even though at this step the receiver may already have determined that he/she understands how to perform the request and is capable of performing it, it must still be determined whether the receiver has the authority to perform the request. This step therefore measures whether the receiver, as part of his or her duty, has the authority to provide the requester with the requested action or information. If the receiver has the necessary authority, the 'yes' option is taken, otherwise the 'no' option is selected. This is the last step for the receiver to measure him- or herself — the next two steps focus on the request itself.

7.3.6 Is the requested action or information available to the public?

This is the first step that may lead to a request being performed, as it is here determined whether the requested action or information is already publicly available. In the case where the request involves an action, it must be determined whether this requested action is available to the public or whether the requester should possess any level of authority to request the specific action. In the case where the requested action is available to everyone, the 'yes' option is taken and the request is performed. If any type of authority is required for the specific action, the 'no' option is chosen and the receiver proceeds to the next step. The procedure is similar for requested information, since it needs to be measured whether the information is public or private. The receiver should have a clear understanding of what information is readily accessible to the public. For example, information in the public domain

for an institution could include contact details and working hours, which could be available on the institution's website and may thus be provided to a requester.

7.3.7 Is this a preapproved request that can be performed to avoid a life-threatening emergency?

The receiver needs to assess the urgency of the request and the consequences for the requester if the request is not performed. If the request does not constitute a life-threatening emergency, the 'no' option must be taken. An example of a life-threatening emergency is where an individual's medical insurance number is required because he/she was injured in an accident. This is a very difficult step to measure, because a skilled social engineer can use a life-threatening emergency to get the receiver to perform an unauthorised request. It should be noted that a request must be performed when an individual's life could be at risk upon denial of the request and could potentially cause harm to the individual. For this specific step, the receiver must be able to clearly determine what constitutes a life-threatening emergency and which requests may be performed during a life-threatening emergency. If the list of requests that may be performed is continually revised, the social engineer will be limited and thus it can be deemed safe to provide the limited set of requests during a life-threatening emergency. In the case where it is known that an organisation does not deal with life-threatening emergencies, this step can be omitted from the model.

7.3.8 Are any of these conditions for refusal true?

In this step, it is determined whether there are any conditions that can constitute a refusal to perform the request. Only four conditions are provided in the model, therefore they do not constitute an exhaustive list. These conditions can be developed to cater for specific organisations, depending on where the model is implemented. The four conditions that are provided are of a very generalised nature to cater for the most probable conditions without mentioning the specific requirement. If any condition here constitutes a sufficient reason for refusal, the 'yes' option is taken; otherwise, the request is performed. The following subsections elaborate on the conditions used in the refusal component of the model.

7.3.8.1 Are there any administrative reasons for refusal?

This condition refers to all the possible administrative reasons that may constitute a sufficient reason for refusal. For example, the requester may be required to provide information if there is an administrative process in place that requires a specific level of authority for requesting the receiver to perform the request. The term 'administrative reason' is used to refer to a wide variety of reasons and keeps the model more generic.

7.3.8.2 Are there any procedural reasons for refusal?

This condition refers to all the possible procedural reasons that may constitute a sufficient reason for refusal. For example, the receiver may have stumbled on a portable storage medium and wants to return it to its rightful owner. However, there is a procedural policy in place that forbids the receiver to plug in any storage medium into a workstation at the organisation and this will then constitute a valid reason to refuse the request and take the 'yes' option. The term 'procedural reason' is used to refer to a wide variety of reasons and it keeps the model more generic.

7.3.8.3 Is this an unusual or new type of request?

One also needs to consider the case where the type of request has not yet been defined in an organisation or where the receiver has never before dealt with such a type of request. In this case, the unusual or novel nature of the request will be sufficient reason to refuse it and take the 'yes' option to further ascertain the requester's identity. For example, the receiver may receive a request to perform a password reset for a colleague; however, the request is received via e-mail and not telephonically as it is usually done. In this case, it is sufficient reason for the receiver to refuse the request.

7.3.8.4 Are there any other reasons for refusal?

This condition is open-ended as the receiver may feel uneasy with the request or there may be a reason why the receiver intuitively does not want to perform the request. Although it is added as an additional safeguard to the model, this step is not seen as redundant, as it provides the receiver with an additional

means to check out information about the requester. Because the receiver will have the ability to further verify the requester in the steps that follow, it may put the receiver at ease to perform the request.

7.3.9 Is the requester's identity verifiable?

The receiver now needs to verify the identity of the requester to be able to make an informed and rational decision about whether the request should be performed. If at this stage of the model the requester's identity cannot be verified, the 'no' option is taken and the request is either deferred or referred. In the case where the receiver can perform steps to identify the requester and the associated roles, authority or additional details, the 'yes' option is chosen to determine the number of levels on which the requester can be verified.

7.3.10 How many verification requirements hold?

Depending on the extent to which the requester's identity can be verified, a different set of states will be examined to determine whether the request should be performed. Important to remember is that the social engineer may be portraying him- or herself as an authority figure in the institution; a computer technician, or any other persona that may elicit compliance. As people, we are inclined to make quick assumptions regarding others and their status, sometimes even based on trivialities such as clothing. If someone is dressed in the proper attire, uses the appropriate institutional jargon or uses an important individual's name, it does not necessarily indicate that such individual is trustworthy. The same holds for physical objects, as individuals are inclined to help other people. If an individual finds a storage medium lying around and the storage medium is marked to be important, the individual will likely feel the urge to either return the storage medium to its rightful owner or curiosity may drive the individual to examine the contents on the storage medium. If the receiver feels unsure at any time, a super user should be contacted to obtain the authority to provide or decline the request.

The following verification requirements should be taken into account and used as a basis for a decision on whether to perform or not perform the request: authority, credibility, previous interaction, and knowledge of the person's existence. The model has only four verification requirements and hence does not constitute an exhaustive list — additional requirements can be added to cater for specialised

environments and specific to particular organisational contexts. The current list of requirements is very broad and only includes the most common verification requirements.

For example, some of the techniques that can aid in the verification of an individual's identity, specifically in a call centre environment, are the following: Caller Identification; Calling back the requester on a predetermined phone number; Requesting an organisational email address; Requesting a secure password; Requesting face-to-face interaction with the individual where proper identification can be provided; Having another employee to vouch for the requester; Contacting the requester's immediate supervisor to verify the former's identity; Using an employee directory [5]. This illustrates how the verification requirements can be changed based on the environment and organisation where the model is implemented.

It is not always possible to verify all of the requirements, so the model has three different outcomes for this specific step. If none of the verification requirements hold, the 'none' option is selected and the request is either deferred or referred. In the case where only one to two requirements hold, the path to further verify the information from a third party source is selected. Lastly, if three or more requirements hold, the receiver proceeds to the final step of the model where it is measured whether the requester has the necessary authority to request either the action or the information. The strictness of this step can also be changed depending on the environment or organisation where this model is implemented. The strictness of this step is highly dependent on the verification requirements that are utilised. Each of the verification qualities is addressed in the next subsection.

7.3.10.1 Can the authority level of the requester be verified?

Authority is an integral part of any institution, with an almost conditioned response from employees to adhere to an authoritative requester's wishes and demands, combined with a fear of punishment should the receiver appear to undermine the requester [11]. For these reasons, impersonating an authoritative individual is a very effective technique used by social engineers to obtain privileged information or actions. The institution needs to provide an environment in which the employee feels comfortable and is indeed expected to question the authority figure's identity before disclosing sensitive information. The employee also needs to know — with the help of a clear institutional policy — on what authorisation

level a particular person of authority is rated, in regard to what privileged information or action can be provided.

The same situation applies in our daily lives, where individuals adhere to rules and regulations put in place by an authoritative figure. Examples of such official procedures are road rules and regulations, and individuals tend to follow these rules and regulations because they create a safe environment for everyone to travel in. Theft is also frowned upon. When a storage medium is found by an individual, the individual will rather want to return this storage medium to its rightful owner, otherwise, it might be considered as stealing the device.

7.3.10.2 Can the credibility of the requester be verified?

The employee needs to judge the level of credibility of the requester. However, this is a challenging task, as establishing credibility is the first aim that the social engineer tries to achieve, and what the attack will be based on. If the requester knows the jargon used by a particular institution, people easily assume that he/she is an employee at their particular institution. The requester could, for example, be an ex-employee who is (still) quite knowledgeable about the jargon and procedures. Such an aggrieved ex-employee may try to seek revenge with the goal of obtaining particular sensitive information. The credibility of the requester is measured on the basis of how well or bad he/she responds to a predefined of set of questions used to determine the credibility of a requester.

Similarly, for objects and items, the credibility of an item has to be measured in terms of whether it conforms to the institution's guidelines. One must examine whether it is the same brand used by the institution or whether the institution's lanyard is attached to it. These are all minor techniques that can be used to ensure that the object or item indeed belongs to or is part of the institution.

7.3.10.3 Did you have a previous interaction with the requester?

If the individual has had previous interaction with the requester, especially a long-standing history of interaction, the decision whether information can be provided will be an easier task. However, limited interactions with the requester, especially by telephone and email alone, should be taken into account in conjunction with other verification techniques, to be able to make an informed and safe decision

regarding the request. This test can only be performed on individuals and not on items or objects that have an inherent request.

7.3.10.4 Are you aware of the existence of the requester?

This refers to the knowledge that the requester exists in the institution or the fact that an outside collaborating partner on a project can support the verification of the requester. However, this should also be used in conjunction with the other verification techniques, as the requester could be a social engineer portraying him- or herself to be some well-known individual in order to get the receiver to perform the request. This awareness test can only be performed on individuals and not when the requester is an object that has an inherent request.

7.3.11 Can you verify the requester through a third party source?

When the requester did not adhere to enough verification requirements and only some of the verification requirements held, this step is reached. It tests whether it is possible to verify the information obtained from the requester through an external source. An example of such a case is where the requester provides information and claims to be part of a specific organisation, as well as to have been requested by his or her organisation to ask the receiver to perform a request. If it is possible for the receiver to contact another individual at this organisation to verify this information, the 'yes' option is chosen, otherwise the 'no' option is taken and the requested is either deferred or referred.

7.3.12 Does the verification process reflect the same information as the verification requirements?

In this step, the receiver will verify all the verification requirements as obtained from the third party source. Continuing with the previous example, the receiver will attempt to contact the organisation and verify the information received from the requester. If the information matches what has been received from the requester, the 'yes' option is taken, otherwise, the request is either deferred or referred.

7.3.13 Does the requester have the necessary authority to request the action or the information?

With the aid of the previous steps the receiver has acquired the necessary knowledge regarding the requester's identity and authority level, together with whether the receiver may perform the request. The receiver can now determine whether the requester has a level of authority on the same level or higher than the level of sensitivity of the request. If the requester has the authorisation on the same authority level as or higher than needed for the particular request, the request is performed. However, if the requester does not have the necessary authorisation, the 'no' option is chosen and the request is deferred or referred.

7.3.14 Defer or refer request

This is the negative result state of the model. In this state the request can either be deferred or handled at a later stage. Deferring the request can also lead to the request never being performed and can be considered as the request having been halted. In the case where the receiver is part of an organisation, there is the option to refer the request to a more authoritative person in the same organisation. This will allow someone else who may have better judgement on whether to perform or halt the request to actually deal with the request.

7.3.15 Perform the request

This is the positive result state of the model. In this state the receiver is allowed to perform the single request from the requester.

7.4 MAPPING SOCIAL ENGINEERING EXAMPLES TO SEADM

Three example scenarios that are provided in this section are used to test the model. In previous work, social engineering was divided into three distinct categories based on the type of communication utilised (see Chapter 2). The three categories are respectively bidirectional communication, unidirectional

communication and indirect communication. An example from each of these categories is used to illustrate how the model works.

In the first scenario, from the bidirectional communication category (see Section 6.3.4), the social engineer attempts to obtain sensitive information that is accessible only to the employees of the organisation. In the second scenario, from the unidirectional communication category (see Section 6.3.7), the social engineer attempts to obtain financial gain by sending out paper mail in which the letter requests a group of individuals to make a small deposit into a bank account owned by the attacker. In the third scenario, from the indirect communication category (see Section 6.3.8), the social engineer attempts to gain unauthorised access to a workstation in an organisation by using a storage medium device.

7.4.1 Scenario One - Bidirectional Communication

In this scenario, as extrapolated from Section 6.3.4, a social engineer attempts to obtain the sensitive information of an organisation to which only the employees of the organisation have access. The information is not available to members of the public. This attack is performed using bidirectional communication. The social engineer sends an e-mail to an employee at the targeted organisation. This e-mail would therefore come from a different e-mail address than the company's own e-mail address. The social engineer uses the friendship and liking compliance principle to persuade the receiver to perform the requested action of providing the information to the social engineer. The attacker may use a pretext such as that he is coming to the organisation for an interview and requires information about the company policies to ensure that he is well prepared for the interview. The rest of this section maps 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.

Do you understand how to perform the request?: The social engineer would have made certain that the targeted employee fully understands the request, is capable of performing the request and has the

authority to perform the request. This will allow the current step, and the following two steps to take the 'yes' option.

Are you capable of performing or providing the request?: As indicated earlier, the 'yes' option is chosen.

Do you have the authority to perform the request?: As indicated earlier, the 'yes' option is taken.

Is the requested action or information available to the public?: In the scenario it states that the information is not available to the public and thus the 'no' option is chosen.

Is this a preapproved request that can be performed to avoid a life-threatening emergency?: This is not a life-threatening request and thus the 'no' option is selected.

Are any of these conditions for refusal true?: Seeing that the requested information is privileged and accessible to employees only, there is an administrative reason for refusal and thus the 'yes' option is selected.

Is the requester's identity verifiable?: In this case, bidirectional communication is utilised; thus it allows for the receiver to communicate back via e-mail and ask more questions to verify the requester. Hence the 'yes' option is taken.

How many verification requirements hold?: In this case, the friendship and liking compliance principle is utilised and the social engineer builds up a trust relationship, via e-mail, with the receiver. The pretext utilised during this attack is that sensitive information about the company policies is requested because the attacker, as part of the pretext, will be attending an interview at the targeted employee's organisation. The receiver is able to verify all four of the verification requirements, i.e. the attacker's authority, credibility, previous interaction and knowledge of existence, as there have been several e-mails before the request. This allows the receiver to choose the 'three or more' option and proceed to the following step.

Does the requester have the necessary authority to request the action or the information?: At this step this attack will fail, even though the social engineer has built up a trust relationship with the receiver, because the attacker is not part of the organisation and thus does not have the necessary authority. The authority level is verified in the previous step, and it has been verified that the authority level of the requester is that of an individual who does not work at the organisation. This will force the receiver to defer or refer the request. The 'no' must be selected here and hence the social engineering attack is thwarted.

7.4.2 Scenario Two - Unidirectional Communication

In this scenario, as extrapolated from Section 6.3.7, a social engineer attempts to obtain financial gain by sending out paper mail. In the letter, a group of individuals are requested to make a small deposit into a bank account owned by the attacker. In this example, the attacker will develop a phishing letter that masks the attacker as a charity organisation requesting donations. The phishing letter contains the contact details, the logo and the purpose of the charity to improve the authenticity of the letter. This attack uses unidirectional communication and thus the receiver is not able to communicate with the attacker. The rest of this section maps the example to the model.

Do you understand what is requested?: The letter from the social engineer should clearly state that a receiver is requested to make a donation to the specific charity. The letter will include all the required details because this receiver cannot communicate with the social engineer. The 'yes' option is taken.

Do you understand how to perform the request?: The social engineer would have ensured that the targeted individual fully understands the request, is capable of performing the request and has the authority to perform the request. This will cause the receiver to select the 'yes' option in this step, as well as in the following two steps.

Are you capable of performing or providing the request?: As indicated before, the 'yes' option is taken.

Do you have the authority to perform the request?: As was the case earlier, the 'yes' option is chosen.

Is the requested action or information available to the public?: The requested action is to make a deposit into the bank account of the requester. This request is directed at the receiver and not at the public. The action of the specific receiver making a deposit is only available to the specific receiver, thus the 'no' option is taken.

Is this a preapproved request that can be performed to avoid a life-threatening emergency?: This is not a life-threatening request and thus the 'no' option is selected.

Are any of these conditions for refusal true?: This request can be seen as either unusual or new as the requester would not usually receive this specific type of letter from the charity. It can also be the case that the requester feels uneasy about the request and his or her uneasiness about the request can be seen as a reason to refuse at this point. The 'yes' option is selected because there is sufficient reason to refuse the request without even verifying the identity of the requester.

Is the requester's identity verifiable?: Since unidirectional communication is utilised in this case, the receiver can only verify the identity using the information as provided in the letter. At this point one can defer or refer the request if it does not contain additional information such as the requester's contact details. In the current scenario, the letter actually contains the contact details of the charity organisation and thus the 'yes' option is chosen.

How many verification requirements hold?: The requirement that the receiver should be aware of the existence of the requester will definitely hold, because the social engineer would have chosen a well-known charity. One can also argue that receiver may have had a previous interaction with the charity; however, from the letter alone, the authority and credibility of the requester cannot be verified. In this case the 'one to two' option is selected.

Can you verify the requester through a third party source?: The receiver will now have the ability to verify the information in the letter directly from the charity organisation. The receiver will make a phone call to the charity to verify the information. It is assumed that the charity organisation can be reached to verify the information and thus the 'yes' option is taken.

Does the verification process reflect the same information as the verification requirements?: It is at this step that the receiver will be able to ask the organisation whether such a letter has in fact been sent out. The charity organisation will deny this and thus the verification process will show that the information provided is not the same as the verification requirements. Consequently, the 'no' option will be taken and the social engineering attack will be thwarted.

7.4.3 Scenario Three - Indirect Communication

In this scenario, as extrapolated from Section 6.3.8, the social engineer attempts to gain unauthorised access to a workstation in an organisation by using a storage medium device. The organisation does not have a company policy in place that disallows employees plugging storage devices into their workstations. The social engineer will leave the device outside the organisation's building to be found by an employee. The device will be infected with a trojan so that when it is plugged into the workstation, it opens a backdoor for the social engineer to connect to the system remotely. As the storage device is left unattended, this attack utilises indirect communication. The rest of this section maps this example to the model.

Do you understand what is requested?: The storage medium device planted by the social engineer should be marked clearly to indicate that it contains important and confidential information. Thus the receiver who finds this device will want to return it to its rightful owner. As it is an inherent request that the receiver should return the device, the request is easily understandable and the 'yes' option is selected.

Do you understand how to perform the request?: The social engineer would have made certain that the storage medium device is deployed at such a location that only individuals who have access to a workstation and who understand how such devices work should find the device. This will cause the receiver to take the 'yes' option in this step as well as in the following step..

Are you capable of performing or providing the request?: As was the case previously, the 'yes' option is selected.

Do you have the authority to perform the request?: In this step, the receiver should ask him- or

herself whether he/she has the authority to plug the storage device into a workstation at the organisation. If there is a company policy that disallows or forbids this, then the 'no' option will be selected and the attack be thwarted. However, for the present scenario there are no company policies in place and thus the receiver has the necessary authority. Consequently, the 'yes' option is taken.

Is the requested action or information available to the public?: The inherent requested action is to return the storage device to its rightful owner. This request is directed at the receiver who found the device. Because only the receiver can perform this action, the 'no' option is taken.

Is this a preapproved request that can be performed to avoid a life-threatening emergency?: The scenario does not involve a life-threatening request and thus the 'no' option is chosen.

Are any of these conditions for refusal true?: In this scenario, the storage device has been marked as confidential and important. Hence, the receiver will not be allowed to plug the device into a workstation. This will be considered as a reason for refusal and cause the 'yes' option to be taken. Administrative and procedural reasons are ruled out for this example because there are no company policies that govern storage devices.

Is the requester's identity verifiable?: Since indirect communication is utilised in this case, the only piece of information the receiver has is the physical storage medium device. Due to the confidentiality of the device, the receiver is unable to verify the requester's identity, therefore the request is deferred or referred and the attack is thwarted. In the present scenario, the request will most likely be referred to another individual in the organisation who is allowed to safely, and on a secure workstation, verify the contents of the storage device and potentially contact the rightful owner.

7.5 CONCLUSION

The protection of information is extremely important in modern society and even though the security around information is continuously improving, a weak point is still the human being who is susceptible to manipulation techniques. This chapter explored social engineering as a domain and social engineering attack detection techniques as a process inside this domain. The first iteration of the SEADM was revisited and significantly improved upon.

The author found that the previous SEADM catered only for social engineering attacks utilising bidirectional communication. This chapter proposed a revised version of SEADM, which caters for all three categories of social engineering attacks. The revised SEADM has been verified using generalised social engineering attack examples. It was shown that the revised SEADM is able to thwart social engineering attacks that make use of either bidirectional communication, unidirectional communication or indirect communication.

The proposed revised SEADM can now be used as a tool to protect oneself against social engineering attacks. Even if the model is not adhered to in respect of every request, it will cause one to think differently about requests — and this is already a huge step in the right direction.

It is important to note, this is not the final iteration of the SEADM. It has been found that the SEADM is extremely effective at detecting social engineering attacks, however, it is not easily extensible. The following chapter attempts to alleviate this problem by generalising the SEADM into a finite state machine.

CHAPTER 8 FINITE STATE MACHINE OF THE SEADM

8.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with the third, and final, iteration of the Social Engineering Attack Detection Model (SEADM). This chapter is based on both a conference paper entitled "Underlying finite state machine for the social engineering attack detection model" and a journal paper entitled "Finite state machine for the social engineering attack detection model: SEADM" which has previously been published by the author [124, 125]. The conference paper has been peer reviewed and was published at the Information Security for South Africa conference in 2017 [124]. The journal paper was a significant extension of the conference paper and was published in the SAIEE Africa Research Journal in 2018 [125]. These articles revisited both the first version of the SEADM, which was proposed in 2010, and the second version of the SEADM, which was proposed in 2015.

The second iteration of the SEADM was effective against detecting social engineering attacks, however, it was not developed with a focus on extensibility. The second iteration could be extended by adding validation questions to the model, however, it was not fully clear on where in the model these questions should be added. The final iteration, which is provided in this chapter, attempted to address this problem by converting the SEADM into a finite state machine. The conversion of the SEADM into a finite state machine replaces the qualitative sub-procedures, of the second iteration, with generalised states that better define the role of each sub-process. This allows one to both add or remove validation questions which directly related to the defined role of each state. This chapter focuses on the final iteration of the SEADM and validates the model by means of social engineering attack examples. The

finite state machine is also validated by proving that the model is both deterministic and correct for all possible input alphabets.

8.2 INTRODUCTION

The previous iteration of the SEADM, SEADMv2, focused on covering all three different types of communication mediums for social engineering attacks, as shown in Chapter 7. Whilst using the SEADM to determine whether it is effective to detect social engineering attacks, it was noted that each set of questions focuses on a specific context. Also, the current iteration of the SEADM does make mention that additional questions can be added to the model to address different implementation environments, but it is not explicitly stated how this should be done.

This chapter focuses on addressing this problem by formalising the latest iteration of the SEADM into a deterministic finite state automata. The author is not aware of similar approaches by other researchers. In its original form, SEADM was constructed as a non-deterministic flow chart that relied on general, qualitative sub-procedures to provide a model for detecting social engineering attacks. While effective as a procedure for reducing risk, the model made no provision and provided no guidance on how additional actions relevant in specific contexts and domains could be included, and at what points in the model these inclusions should be placed. Due to the inclusion of cycles in the SEADM model, the process was also non-deterministic, which added additional and unnecessary complexity in implementing the process.

This research aims to improve the extensibility of the SEADM, and to reduce its implementation complexity by restructuring the process to be cycle-free and deterministic. The extensibility of the model is addressed by replacing the qualitative sub-procedures with generalised states that better define the role of each sub-process, while treating questions posed in each state as general examples that may be expanded or removed, and not as a definitive collection of necessary queries. Organising the model as a finite set of generalised states provides a more concise representation of the process that encapsulates the broad set of questions into distinct units of related, deterministic, context specific states. This adjustment is intended to improve extensibilty, while simultaneously reducing the complexity of implementing the model as an organisational process or in software.

The remainder of the chapter is constructed as follows. Section 8.3 proposes the underlying deterministic finite state machine of the SEADM. Section 8.4 provides a discussion on each of the states on how they were derived from the SEADM. Section 8.5 concludes this chapter.

8.3 UNDERLYING FINITE STATE MACHINE OF THE SEADM

A finite state machine (also known as finite state automaton) is an imaginary machine that embodies the idea of a sequential circuit. It has a finite set of states with a start state and accepting states, and a set of state transitions [126]. Finite state machines are commonly employed in the design and implementation of modern software and electronics, and range from simple and highly abstract models of computation or processing to complex and concrete executable mechanisms and physical circuitry.

Finite state machines can be deterministic or non-deterministic. A deterministic machine has exactly one path for every input-state pair. In a non-deterministic machine there may be multiple valid transitions for every input-state pair, and the chosen transition is not defined; any transition can be followed. A deterministic finite state machine is a state machine that is guaranteed to complete for all inputs in a finite amount of time, while a non-deterministic finite state machine may execute indefinitely or fail to progress toward completion given a certain set of inputs. A finite state machine is provably deterministic if and only if it is both free of cycles (that is, no state is ever revisited after being processed once) and defines a transition to a new state for each potential input in every state (that is, any valid input into a state results in a transition to a new state). These two properties together preclude the possibility of the state machine entering an infinite loop, either within a single state or between a collection of states, thereby ensuring that processing will always complete within a finite number of steps.

The finite state automaton described in this chapter is an abstract or general model, and is intended to provide a structured but flexible deterministic high-level model of the steps taken to mitigate a social engineering attack, improving upon the original SEADM flow-chart model. While the SEADM flow-chart provides a static procedural template for implementing detection mechanisms for social engineering attacks, the abstract state diagram provides a more accessible, abstract and extensible model that highlights the inter-connections between task categories associated with different scenarios. The abstract state-based model is intended to help facilitate the incorporation of domain, system or

organisation specific extensions by grouping related activities into distinct categories, subdivided into one or more nodes. Should a specific task, necessary in a particular domain, systems or organisational context, not be included in the flow-chart, the state diagram may be used to identify the correct location within the model to incorporate the task. In addition, it facilitates additional analysis on state transitions that are difficult to extract from the more verbose flow-chart.

The current iteration of the SEADM, as depicted in Figure 7.1 and repeated here as Figure 8.1, utilised four different state categories: the request, receiver, requester and third party. The request states, indicated in yellow, directly deal with information about the request itself. The receiver state, indicated in blue, directly deals with the person handling the request and whether this person (the receiver) understands and is allowed to perform the request. The requester states, indicated in green, directly deal with the requester and whether any information about the requester can be verified. The third party states, indicated in red, consider the involvement of a third party in the model and whether the information about the requester can be externally verified.

The same four categories and colour schemas are maintained in the state machine as they depict the primary topic that each specific state deals with, allowing one to better understand the state machine. The state machine is depicted in Figure 8.2. Each state has an associated letter which explains which condition needs to be met before the transition can be performed. As an example, a state can have an alphabet of Y and $\neg Y$. The symbol $\neg$ indicates negation, so $\neg Y$ is Not Y, or more accurately, the opposite of Y.

The states in Figure 8.2 are explained as follows:

- S_1 deals with understanding the request. The request is either 'understood' (U) or 'not understood' ($\neg U$) by the receiver.

- S_2 deals with requesting more information in an effort to understand the request. There is either 'sufficient information' (I) or 'insufficient information' ($\neg I$) for the receiver to perform the request safely.

- S_3 deals with the capability of the receiver to perform the request. The receiver is either 'capable' (C) or 'incapable' ($\neg C$) of performing the request.

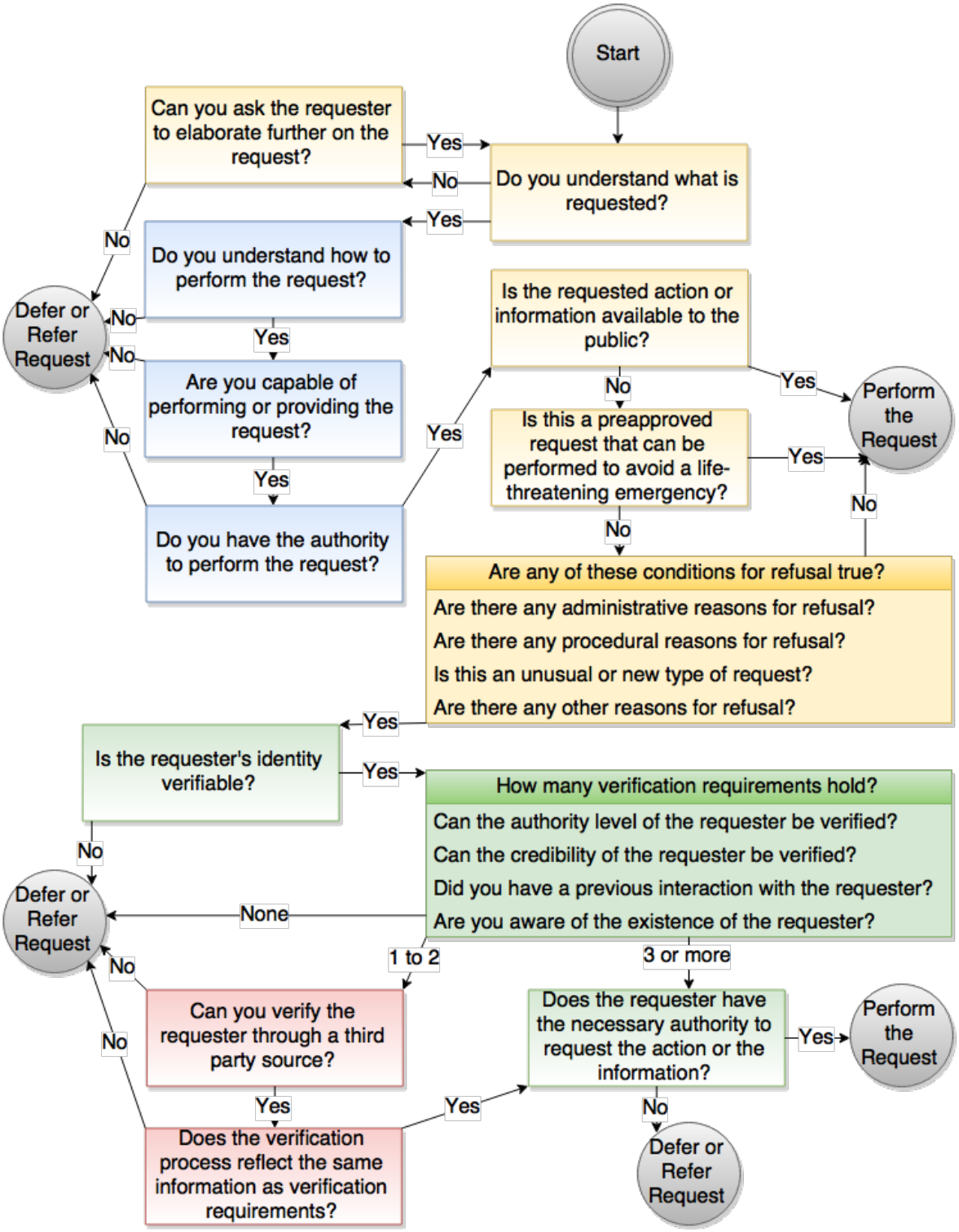

Figure 8.1: Social Engineering Attack Detection Model

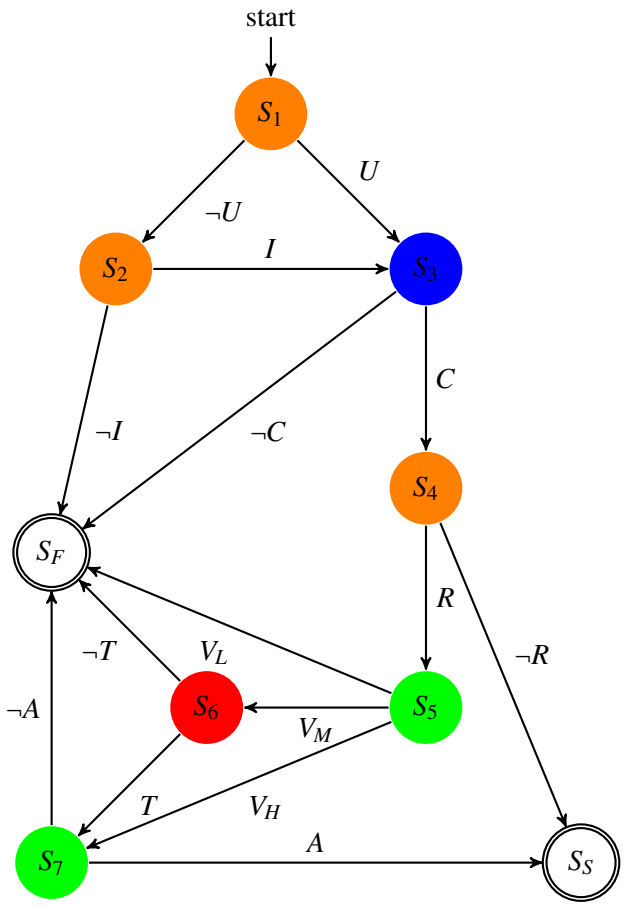

Figure 8.2: Underlying Finite State Machine of the SEADM

- S_4 deals with further verification requirements that may need to be met. The request either has further 'verification requirements' (R) or has 'no verification requirements' ($\neg R$).

- S_5 deals with whether the receiver can verify and trust the identity of the requester, and how many of the verification steps hold. The verification steps hold to either a 'high amount' (V_H), 'medium amount' (V_M) or 'low amount' (V_L) which governs how the receiver should proceed.

- S_6 deals with third party verification. The requester is either 'verified' (T) by a third party or is 'not verified' ($\neg T$).

- S_7 deals with authority. The requester either has 'sufficient authority' (A) or 'insufficient authority' ($\neg A$) for the particular request.

- S_F is an end state. This state indicates the failure state. The request should not be performed by the receiver.

- S_S is an end state. This state indicates the success state. The request should be performed by the receiver.

The states are elaborated further in section 8.4. A description of the full state machine in mathematical notation follows. The finite state machine is a 5-tuple consisting of the finite set of input alphabet characters Σ, the finite set of states Q, the start state S_0, a set of accepting states F, and a set of state transitions δ that contains 3-tuples representing state transitions. A 3-tuple in δ consists of a current state, a current input and the next state.

$$
\begin{aligned}
\Sigma &= \{U, \neg U, I, \neg I, C, \neg C, R, \neg R, V_L, V_M, V_H, T, \neg T, A, \neg A\} \\
Q &= \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_S, S_F\} \\
S_0 &= S_1 \\
\delta &= \{ \\
&\quad (S_1, U, S_3), (S_1, \neg U, S_2) \\
&\quad (S_2, I, S_3), (S_2, \neg I, S_F) \\
&\quad (S_3, C, S_4), (S_3, \neg C, S_F) \\
&\quad (S_4, R, S_5), (S_4, \neg R, S_S) \\
&\quad (S_5, V_L, S_F), (S_5, V_M, S_6), (S_5, V_H, S_7) \\
&\quad (S_6, T, S_7), (S_6, \neg T, S_F) \\
&\quad (S_7, A, S_S), (S_7, \neg A, S_F) \\
&\quad \} \\
F &= \{S_S, S_F\}
\end{aligned}
$$

Using both Figure 8.2 and the provided mathematical model it is straightforward to infer a state

transition table. Table 8.1 depicts all the possible state transitions given a specific input for each state. For all input states, the output is either a terminal node or a node with a higher state index. This illustrates that the state machine is deterministic, eliminating cycles present in the original SEADM flowchart.

Table 8.1: State Transition Table for the SEADM

Input \ State	S_1	S_2	S_3	S_4	S_5	S_6	S_7
U	S_3	—	—	—	—	—	—
$\neg U$	S_2	—	—	—	—	—	—
I	—	S_3	—	—	—	—	—
$\neg I$	—	S_F	—	—	—	—	—
C	—	—	S_4	—	—	—	—
$\neg C$	—	—	S_F	—	—	—	—
R	—	—	—	S_5	—	—	—
$\neg R$	—	—	—	S_S	—	—	—
V_L	—	—	—	—	S_F	—	—
V_M	—	—	—	—	S_6	—	—
V_H	—	—	—	—	S_7	—	—
T	—	—	—	—	—	S_7	—
$\neg T$	—	—	—	—	—	S_F	—
A	—	—	—	—	—	—	S_S
$\neg A$	—	—	—	—	—	—	S_F

To further show that the state machine model is deterministic, resulting in a valid outcome of either success or failure for all given alphabet sequences, a transition table indicating all possible input alphabet sequences (paths) and their corresponding results are shown in Table 8.2. Each row in the table represents a path. Σ_i indicates the input character is the i-th character in the path. The symbol $\forall$ indicates no transition occurred in the i-th position of the path in the case of shorter paths. This table shows that for all possible paths, the state machine returns either success or failure in a finite number of steps, and is thus deterministic.

Having considered the high-level state-based model for the SEADM, the following section elaborates on the purpose of each state and exactly how it was derived from the SEADM.

Table 8.2: State Transition Table for all Input Alphabets

	Input Alphabet							Output	
No	Σ_1	Σ_2	Σ_3	Σ_4	Σ_5	Σ_6	Σ_7	S_S	S_F
1	U	$\forall$	C	R	V_H	$\forall$	A	✓	—
2	U	$\forall$	C	R	V_M	T	A	✓	—
3	U	$\forall$	C	R	V_M	T	$\neg A$	—	✓
4	U	$\forall$	C	$\neg R$	$\forall$	$\forall$	$\forall$	✓	—
5	U	$\forall$	C	R	V_H	$\forall$	$\neg A$	—	✓
6	U	$\forall$	C	R	V_M	$\neg T$	$\forall$	—	✓
7	U	$\forall$	C	R	V_L	$\forall$	$\forall$	—	✓
8	U	$\forall$	$\neg C$	$\forall$	$\forall$	$\forall$	$\forall$	—	✓
9	$\neg U$	$\neg I$	$\forall$	$\forall$	$\forall$	$\forall$	$\forall$	—	✓
10	$\neg U$	I	C	R	V_H	$\forall$	A	✓	—
11	$\neg U$	I	C	R	V_M	T	A	✓	—
12	$\neg U$	I	C	R	V_M	T	$\neg A$	—	✓
13	$\neg U$	$\forall$	C	$\neg R$	$\forall$	$\forall$	$\forall$	✓	—
14	$\neg U$	I	C	R	V_H	$\forall$	$\neg A$	—	✓
15	$\neg U$	I	C	R	V_M	$\neg T$	$\forall$	—	✓
16	$\neg U$	I	C	R	V_L	$\forall$	$\forall$	—	✓
17	$\neg U$	I	$\neg C$	$\forall$	$\forall$	$\forall$	$\forall$	—	✓

8.4 DISCUSSION OF EACH STATE

This sections explains how each of the states has been designed and integrated from the SEADM model. Throughout this discussion the alphabet of the states are provided. During the discussion on the states, it is also shown how each state relates to the SEADM. Each state has been generalised to such an extent that it can contain any number of questions required to achieve a specific transition result. This provides a rough guide for flexibility and extensibility, depending on the particular context the model is applied to.

8.4.1 State S_1: Understanding the Request

This state considers whether the receiver of the request fully understands the request in its entirety. This means that the requester should have provided all the information required to enable the receiver to perform the request in full. The question that was provided in the SEADM was "Do you understand what is requested?"

In the SEADM there was only one question asked here. This has been created as the start state as it is still important to fully understand what is requested before the request can be processed any further. In this state the alphabet is as follows:

- U represents that the request is understood in its entirety and that the receiver has all the necessary information in order to be able to perform the request.

- $\neg U$ represents that the request is not fully understood and that the receiver requires more information about the request.

This state can only transition in one of two ways and is depicted as follows:

- (S_1, U, S_3)

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

8.4.2 State S_2: Requesting information to fully understand the request

In the SEADM this state was previously grouped together with the previous question. This resulted in a loop, as the receiver could always ask the requester to elaborate further. At some point the requester would no longer be able to elaborate on the request as there would be no more additional information available, and the loop would terminate. At what point the loop would terminate was unclear, and could not be formalised into a deterministic state machine without additional complexity.

Representing this task as a distinct state more clearly defines this, as it deals directly with the information that is required to complete the request and not whether one can request more information. This state considers whether the requester can provide information to such an extent that the receiver is able to fully understand the request. Previously the question that was asked was as follows, "Can you ask the requester to elaborate further on the request?" In this state all questions should be aligned with whether the requested information, in the case that additional information is provided, allows the receiver to understand the request in full or not. In this state the alphabet is as follows:

- I represents that the requester can and has provided enough information for the request to be understood in its entirety by the receiver.

- $\neg I$ represents that the receiver is unable to understand the request in full. This can be because the requester could not provide more information, the requester could not be reached, or that the information provided by the requester was insufficient or incomplete.

This state can only transition in one of two ways and is depicted as follows:

- (S_2, I, S_3)

- $(S_2, \neg I, S_F)$

8.4.3 State S_3: Does the receiver meet the requirements to be capable of performing the request?

State S_3 is used to determine whether the receiver meets all the requirements to perform the request. This state is associated with three questions in the SEADM. The questions are as follows:

- "Do you understand how to perform the request?"

- "Are you capable of performing or providing the request?"

- "Do you have the authority to perform the request?"

The goal of this state is to ensure that the individual who deals with the specific request has the necessary skill level and has the required authority to perform the request. Each question in this state deals directly with the role of the receiver and determines whether the request has been issued to the correct receiver. In this state the alphabet is as follows:

- C represents that the receiver has met all the requirements in order to be capable of performing the requested action or to provide the requested information.

- $\neg C$ represents that the receiver does not meet the requirements to be capable of dealing with the request.

This state can only transition in one of two ways and is depicted as follows:

- (S_3, C, S_4)

- $(S_3, \neg C, S_F)$

8.4.4 State S_4: Does the request have any further requirements that need to be met before the request can be provided?

State S_4 deals with the the request itself and whether there are any special conditions or requirements, such as policies and procedures that need to be followed, associated with the request. Examples of special conditions include whether the request relates to information already in the public domain and is accessible to all or whether the request is a life threatening emergency. If the request is already in the public domain, for instance, there are no further requirements that need to be met. In the event of a life threatening emergency, the outcome depends on whether there are set policies in place. For instance, if there is a policy in place that allows medical personnel to rather err on the side of caution in order to save an individuals life, there will be no further requirements that need to be met and the request may take place. The previous model dealt with these examples using the following questions:

- “Is the requested action or information available to the public?”

- "Is this a pre-approved request that can be performed to avoid a life-threatening emergency?"

This state also deals with any specific requirements that need to be met. Examples of such requirements are any policies or procedures that are in place that require further verification of the requester. This state can also cater for unusual requests. An unusual request is a request that is new to the receiver and is typically something that the receiver does not deal with on a daily basis. By following the rest of the model, the receiver ensures that adequate information about the requester is obtained before the request is performed. It also allows the receiver time to think about the request and whether the request should be performed for the requester. The previous model dealt with these examples using the following questions:

- "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?"

The goal of this state is to ensure that any request which is already public information should be immediately performed and that any request that requires verification should result in further interrogation of the requester. In this state the alphabet is as follows:

- R represents that the request has further verification requirements that need to be met in order to be able to perform the requested action or to provide the requested information.
- $\neg R$ represents that the request has no further verification requirements and that the requested action can be performed or that the requested information can be provided.

This state can only transition in one of two ways and is depicted as follows:

- (S_4, R, S_5)

- $(S_4, \neg R, S_S)$

8.4.5 State S_5: To what extent is the requester's identity verifiable?

State S_5 aims to address the question of the extent of verifiability of the requester's identity. The identity of the requester is determined to verify if the requester has sufficient privileges to request the specific action or information. It is not always possible to verify the identity of the requester in full. The type of communication medium that is utilised by the requester usually will dictate to what extent the identity of the requester can be verified. Typically, if the requester makes the request in person one is able to verify significantly more information about the requester than what one can do over an e-mail or postal mail. It may also be the case that the request is performed over unidirectional communication and that the receiver is unable to receive any further communication from the requester.

The previous model provided a fourth question as to whether the requester's identity is verifiable at all. The state machine has combined not being verifiable and having a low level of verification as the same transition as both states lead to S_F. The questions that were previously used to perform the verification requirements are as follows:

- "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?"

All of the questions catered for a single point of verification. It was also noted in the previous model that the level of verification required to transition to different states should be based on what type of environment the model is applied to. The state machine makes this more generalised by having three possible transitions where there is a low, medium or high level of verification. The threshold for low, medium and high must still be determined based on the environment or context, but the state diagram is more flexible when more questions are added. In this state the alphabet is as follows:

- V_L represents that there is a low level of verification as only a few of the verification elements could be verified.

- V_M represents that there is a medium level of verification as some of the verification elements could be verified, but not all of them.

- V_H represents that there is a high level of verification as most of the verification elements could be verified.

This state can only transition in one of three ways and is depicted as follows:

- (S_5, V_L, S_F)

- (S_5, V_M, S_6)

- (S_5, V_H, S_7)

8.4.6 State S_6: Can you verify the requester's identity from a third party source?

State S_6 is only entered when there is a medium level of verification requirements that have been met. If the requester could not be fully verified directly, the third party source is utilised to determine whether the information provided by the requester was indeed truthful. The previous model did not elaborate much on the third party verification and only had two questions associated to it, as follows:

- "Can you verify the requester through a third party source?"

- "Does the verification process reflect the same information as the verification requirements?"

The questions did not mention specifically which verification requirements needs to be verified. Utilising the state machine one could build intelligence into the model to only ask the questions where the verification requirements were obtained from the requester. This is just one of the ways that the

extensibility of the state machine improves upon the proposed model. In this state the alphabet is as follows:

- T represents that the verification requirements, as obtained from the requester, corresponds to the information that was obtained from the third party source.

- $\neg T$ represents that the verification requirements, as obtained from the requester, does not correspond to the information that was obtained from the third party source.

This state can only transition in one of two ways and is depicted as follows:

- (S_6, T, S_7)

- $(S_6, \neg T, S_F)$

8.4.7 State S_7: Does the authority level of the requester allow the requester sufficient rights to request the action or information?

State S_7 utilises all the information obtained throughout the model and asks the receiver questions based on the information obtained. This state aims to determine whether the requester has the necessary authority and rights to allow the requester to request the action or the information. The previous model only asked "Does the requester have the necessary authority to request the action or the information?" This state elaborates on this by allowing the receiver to verify whether the requester has sufficient authority to gain access to the request or the information. In this state the alphabet is as follows:

- A represents that the requester has sufficient authority to be allowed to request the receiver to perform the action or to provide the information.

- $\neg A$ represents that the requester does not have sufficient authority and is thus not allowed to request the receiver to perform the action or to provide the information.

This state can only transition in one of two ways and is depicted as follows:

- (S_7, A, S_S)
- $(S_7, \neg A, S_F)$

8.4.8 State S_F: Halt the request

This is the negative result state. In this state the request will be halted. In some environments one could opt to rather defer the request to a more authoritative receiver. Deferring the request can also lead to the request never being performed and can be considered as the request having been halted. In the case where the receiver is part of an organisation, there is the option to refer the request to a more authoritative person in the same organisation. This will allow someone else who may be better equipped to determine whether to perform or halt the request.

8.4.9 State S_S: Perform the request

This is the positive result state of the model. In this state the receiver is allowed to perform the single request from the requester.

The following section concludes the chapter with a summary of the advantages of the underlying finite state machine.

8.5 CONCLUSION

This chapter improves on the SEADM by providing the underlying finite state machine which allows researchers to better understand and utilise the SEADM. Representing the SEADM as a finite state machine allows one to have a more concise overview of the process that is followed throughout the model. The model provides a general procedural template for implementing detection mechanisms for social engineering attacks. The state diagram provides a more abstract and extensible model that highlights the inter-connections between task categories associated with different scenarios. This chapter also shows that the finite state machine is both deterministic and correct for all possible input alphabets, simplifying the process of implementing the SEADM model either as a process or

in software. The state diagram is currently implemented as a mobile application as part of a social engineering prevention training tool [127].

The SEADM, with the underlying finite state machine, can be used as a general framework to protect against social engineering attacks. Even if the model is not adhered to in respect of every request, it will cause one to think differently about requests — and this is already a step in the right direction.

The next chapter revisits the research contribution chapters of this thesis by performing a critical evaluation of all of the research performed.

CHAPTER 9 CRITICAL EVALUATION OF THE RESEARCH CONTRIBUTION

9.1 CHAPTER MOTIVATION

The purpose of this chapter is to provide the reader with a critical evaluation of the research contribution in this thesis. This chapter is different from the other contribution chapters, as it is not a chapter which has an already published conference paper or journal article attached to it. This chapter rather takes each of the publications, in the order that they were presented in this thesis, and provides a critical overview of them and what impact they have on the field of social engineering. This critical evaluation allows the author to reflect on the work that has been performed and provides the reader with some insight into the thought process of the author and the research process that was followed.

9.2 INTRODUCTION

The journey from Chapter 2 up to and including Chapter 8 allowed the author to provide the reader with a full overview of all the research that was performed throughout this thesis. The research topic of social engineering attack detection models originated from the author initially attempting to identify a research gap on how the fields of computer security and psychology can have overlapping areas. The first publication into this field, by the author, was the first iteration of the SEADM [44]. This publication piqued the interest of the author into the field of social engineering and ultimately the field of social engineering was chosen as the focal point of this thesis.

Throughout the research process, the author faced many challenges regarding the chosen topic. The initial literature review indicated that even though there are several publications on social engineering,

very few of the studies actually utilise the same definition for the term. It was at this point that the author realised the immediate need to consolidate all the known definitions within the field and to develop a definition that encapsulates all the existing definitions into a singular definition.

The author also realised at an early stage that there is no existing ethical guidelines for when research is performed within the field of social engineering. The author took it upon himself to evaluate the three mainstream ethical perspectives and determine how they can be applied to social engineering research. The research into the field of ethics was an extremely useful avenue as it allowed the author to fully realise what adverse effects social engineering can have on individuals.

The conscious decision was made that all the proposed models can only be tested by using examples from literature, as the ethical constraints indicated that performing the social engineering attacks can potentially be harmful to individuals. It was identified that even though there are several documented examples of social engineering, they did not always have all of the details regarding how the attack was planned, how information gathering was performed, what techniques were used or which compliance principles were used. Having studied several of the documented examples, the author realised that there is a need for a formalised social engineering attack framework (SEAF). The development of the SEAF allowed the author to better comprehend all of the elements that form part of a social engineering attack.

The SEAF also played an integral part in the development of the social engineering attack examples. Since the SEAF was developed based on the social engineering examples in literature, the author was able to extrapolate fully fledged social engineering attack examples based on real-world examples. The social engineering attack examples that were created improves on the examples in literature by both generalising the examples and to ensure that all elements of the social engineering attack is captured. The social engineering attack examples were created in as much detail as possible so that they can be utilised by other researchers, as well as to validate models within the social engineering domain as a whole. The social engineering attack examples allow researchers to evaluate and validate models without having the need to perform social engineering attacks and potentially harming participants of the associated studies. The models can now be fully evaluated and validated without the need of performing any of the attacks.

Having performed all the ground work to be able to evaluate and validate models within the field of

social engineering, the author was able to revisit the initial research topic of the SEADM. The first task was to revisit the initial model which was mostly aimed at a call centre environment and to improve upon this model by developing a model which was able to cater for all three forms of communication mediums utilised within social engineering. This led to the development of the second iteration of the SEADM.

The second iteration of the SEADM was already able to be evaluated and validated using the social engineering attack examples. However, throughout the evaluation and validation, it was discovered that the second SEADM iteration is not easily extensible. This led to the third and final iteration of the SEADM which was converted into a finite state machine. Having the SEADM as a finite state machine allows for the extensibility of the model as each state clearly defines the role of that state and it is intuitive to add or remove qualitative questions from each state. The third iteration of the SEADM only focused on addressing the extensibility of the model and did not hamper the effectiveness of detecting social engineering attacks. The author used the social engineering attack examples to re-evaluate and validate the final iteration of the SEADM and it was found that SEADM was able to effectively assist individuals in detecting social engineering attacks.

The journey from consolidating the definitions regarding social engineering, evaluating the ethical constraints of social engineering research, developing the social engineering attack framework which led to the development of social engineering attack examples and finally revisiting and performing several iterations of the SEADM was an extensive process. The research journey is illustrated in Figure 9.1 to provide the reader a better understanding of the journey as a whole.

Ultimately, this journey has led to the authoring of several publications within the field of social engineering and the author was able to make a large contribution within the field of social engineering. The rest of this chapter is dedicated to critically evaluating each of these contribution topics and how they all ultimately have an impact on the field.

The remainder of the chapter is constructed as follows. Section 9.3 revisits the proposed social engineering definitions and discusses the importance of these definitions. Section 9.4 discusses the ethical perspectives and how they impacted this thesis. Section 9.5 revisits the social engineering attack framework by providing the advantages and disadvantages, whilst section 9.6 provides the reader with the discussion on the social engineering attack examples. Section 9.7 provides a critical evaluation

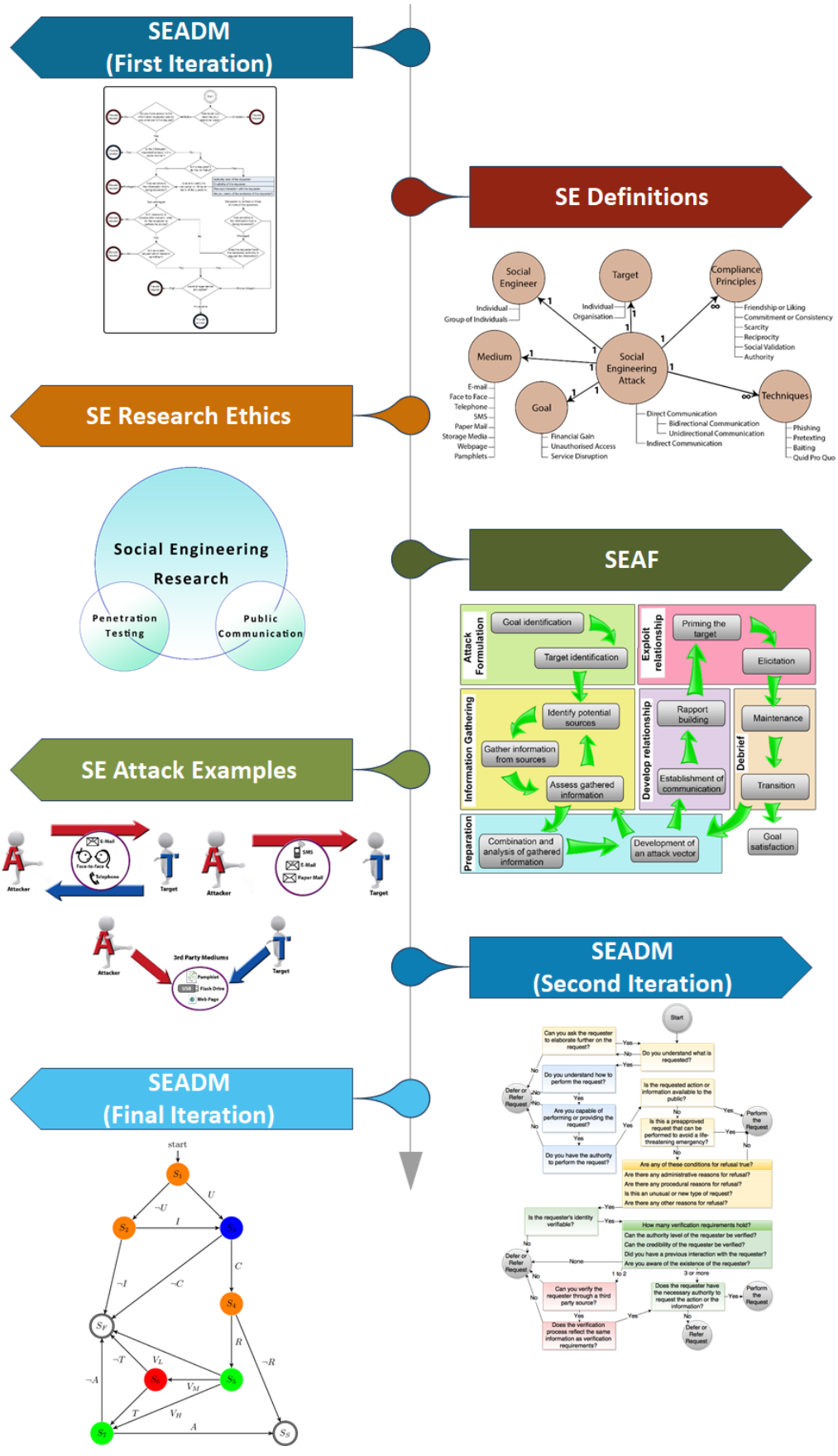

Figure 9.1: Social Engineering Research Journey

of all three iterations of the SEADM and the improvements that each of the iterations brought along. Section 9.8 concludes this chapter by providing a summary of the contributions of this thesis alongside a critical evaluation of these contributions.

9.3 CONSOLIDATED SOCIAL ENGINEERING DEFINITIONS

Chapter 2 allowed the author to introduce the reader to all of the existing definitions for social engineering. The decision was made to consolidate all of the existing definitions of social engineering into a single consolidated one. The consolidated definition for *Social Engineering* is as follows: "The science of using social interaction as a means to persuade an individual or an organisation to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity." Chapter 2 also has proposed definitions for the terms *Social engineer (in both the noun and verb form)* and *Social Engineering attack*.

The benefit of having defined these terms allowed this thesis to have a consistent definition throughout the thesis for the terms. It is also beneficial for other researchers, as consistent definitions allows for research to be comparable to one another. The author also made the decision to ensure that in all the published work, throughout this thesis, mention must clearly be made of the standardised definitions so that any researcher who stumbles upon any of the research articles can make use of the proposed definition.

It should be noted that it is possible for other researchers to disagree with the definition, however, the author is of the opinion that the proposed definition encapsulates all of the existing definitions that were identified in the literature review. Having explored such a vast amount of existing definitions, it also broadened the view of the author and allowed to the author to realise that social engineering can only occur over three communication mediums.

The three communication mediums that were identified was bidirectional communication, unidirectional communication and indirect communication respectively. Having the communication medium form part of the definition also allows the definition to cater for all possible forms of social engineering.

The author also noted that throughout literature the term phishing and social engineering tend to be used inter-changeably, and this is also something that the definition aimed to address. Phishing, albeit one of the most common forms of social engineering, cannot merely replace the term social engineering as it depicts only a subset of the potential social engineering attacks.

The author also depicted social engineering attack as an ontological model. The ontological model is something that can be expanded upon in the future as technology evolves and more mediums, compliance principles and techniques for social engineering becomes available. The ontological model was only developed to aid with the development of the definition and was not further expanded upon within this thesis.

The following section revisits the ethical considerations which guided how the research was performed within this thesis.

9.4 ETHICS REGARDING SOCIAL ENGINEERING RESEARCH

The author had limited knowledge about the ethics research field when the decision was made to explore the field of ethics with regards to social engineering. The decision was made to perform research within the field of ethics and further explore how ethics relate to social engineering research. This research could only be performed by co-authoring with experts within the field who ensured that the research was correct. Having partnered with experts in the field, the author was able to publish both a conference paper and a journal article within the field of ethics [46, 66].

The research into ethics allowed the author to better understand the importance of ethical considerations that needs to be taken into account during research. The thesis explored the three main ethical perspectives namely utilitarianism, deontology and virtue ethics. Each of the ethical perspectives was tested against examples from the public communication, penetration testing and the social engineering research domains.

The research clearly indicated that utilitarianism allows a much larger amount of research to be performed, where as deontology was the most strict ethical perspective by limiting what can be performed. The author is of the opinion that one should rather consider all three ethical perspectives

when performing social engineering research and rather understand why certain items are either acceptable or not.

Having a much clearer understanding as to why something can cause potential harm, can allow the researcher to have a better understanding on how to perform an action in the most ethically correct manner. It should be noted, that different universities subscribe to different ethical perspectives and one should always be guided by the ethical perspective that is prescribed.

The University of Pretoria mostly subscribes to the deontological perspective and due to this, the constraint was put onto this thesis that none of the models can be validated using real-life experiments. This indicated to the author that there is a significant gap in the literature as there is no social engineering attack examples to validate models within the domain. This was addressed by the development of the SEAF and the SEADMs.

The author is of the opinion that it may be an interesting exercise to evaluate the model using real-life experiments, however, the risk of harm to an individual will always be present. This thesis focused on eliminating any possible harm to individuals and thus the model was evaluated using the social engineering attack examples. The fact that the model was only tested using examples does not discredit the results of the testing as the model asks questions which are interpreted by the examples, similar to how a human would respond. There is also no form of controlled environment where the test can be performed, without ever misleading the participant or providing the participant with partial information only.

The following section discusses the social engineering attack framework and how it was used to develop the social engineering attack examples.

9.5 SOCIAL ENGINEERING ATTACK FRAMEWORK

Chapter 5 focused on the proposal of a SEAF. The SEAF was proposed as an extension to Kevin Mitnick's social engineering attack cycle. The figure of the SEAF, figure 5.2 is repeated here for ease of discussion as Figure 9.2.

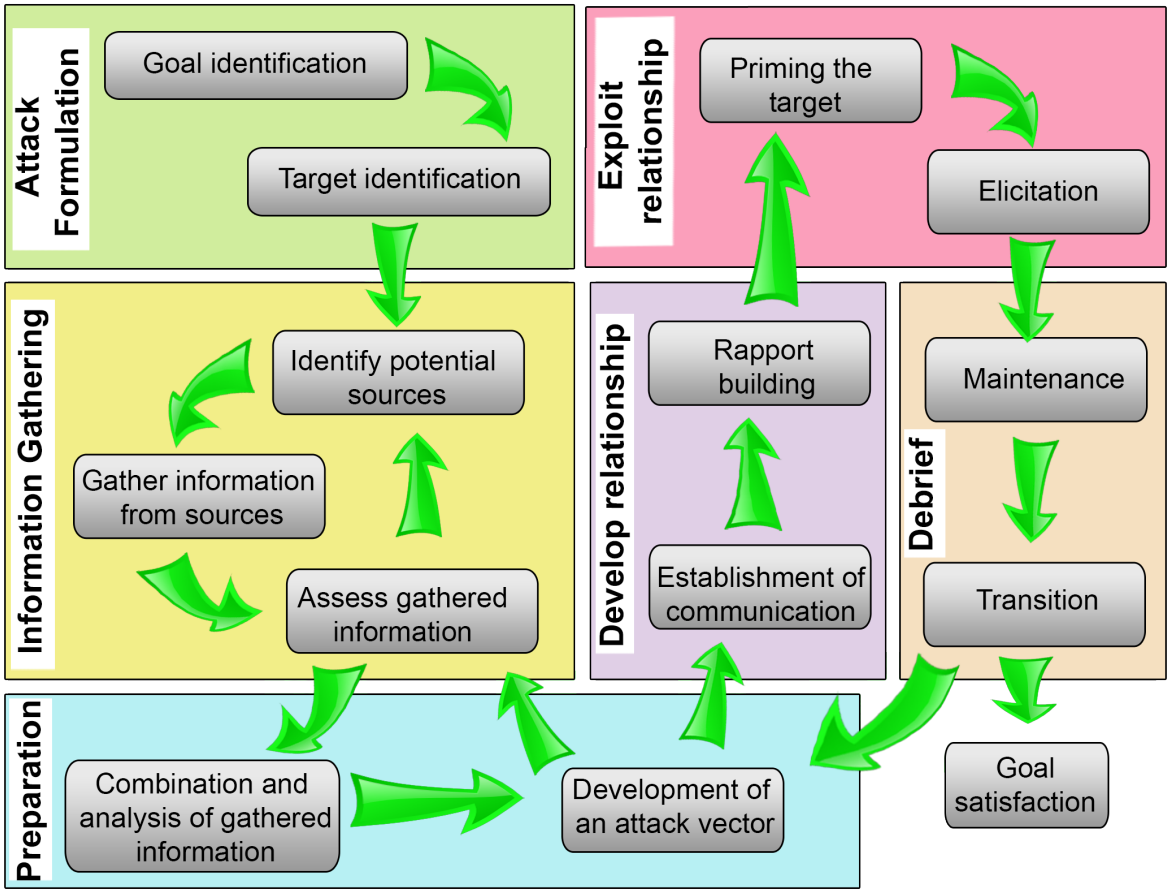

Figure 9.2: Social Engineering Attack Framework (Repeated)

The purpose of the SEAF was to have a framework which can be utilised to both form new social engineering attacks and to document social engineering attack examples from literature. The SEAF caters for all of the phases of a social engineering attack, even in the cases when one has to revisit the information gathering phase. The author is of the opinion that any social engineering attack should be able to fit within the proposed framework. The SEAF is a very useful tool when social engineering attacks are being planned as it allows one to conceptualise all of the elements of the attack, before it is performed.

Having developed the SEAF, the author realised that this framework actually serves a dual purpose and can be utilised both offensively and defensively. The framework can be used defensively in aid of the development of social engineering early warning or detection models. Additionally, it can be used as a training tool to educate people on how social engineering attacks are performed through the simulation of social engineering attack examples. In terms of the offensive approach, this tool can also be used by potential social engineers to better formulate their attacks and to achieve more success. It

is a classic example in computer security that researchers sometimes propose models which can aid attackers, however, one should never shy away from proposing these models as they ultimately aid in detection as well. This thesis has also shown that the SEADM is able to detect all attacks which utilise the SEAF and this is already a defensive mechanism against the malicious use of the SEAF.

The real benefit of the SEAF is how it has aided the author to develop social engineering attack examples from documented social engineering attacks from literature. The development of the social engineering attack examples was the initial driving force behind the development of the SEAF and the critical evaluation of the social engineering attacks is discussed in the following section.

9.6 SOCIAL ENGINEERING ATTACK EXAMPLES

The social engineering attack examples were developed and proposed in Chapter 6. The social engineering attack examples were created by revisiting literature on exisiting social engineering attacks that have been reported on in the past. These examples were grouped into categories, based on the three different types of communication that can be utilised, and it was further grouped into attacks that were of a similar nature. This allowed the author to have at least three documented social engineering attacks, from which information can be extrapolated into a singular generic attack example.

The first step was to map the attack example onto the ontological model of a social engineering attack. This was done to ensure that the scenario has all the required elements of a social engineering attack. Additionally, by mapping it first to the ontological model, it makes the scenario modular in the sense that a single item in the ontological model can be swapped out and it should still be possible to formulate the scenario as a whole. An example of this is where the compliance principle that is used on an attack can be the *authority principle* and this principle can easily be swapped out with the *scarcity principle*. An example of this is where the e-mail can read that a manager demands that you click on a link and this will be changed to where the e-mail reads that the offer will expire if the person does not click on the link. In this scenario, all of the elements of the attack stays similar in nature, the only change will be with how the text in the e-mail is worded.

The second step was to map the populated ontological model, as well as the information from literature on how the attack was performed, onto the SEAF. The SEAF allowed the author to almost write a

narrative on how the attack was performed. The narrative allowed the author to fully delve into the details on how the attack is performed and to further explain how the victim will fall prey to such an attack. The author is also of the opinion that having narrative based social engineering attack examples, allows for easier validation of social engineering models. The narrative approach provided as much information about the examples as possible by writing the examples in a story form. This allows the examples to be easily followed and it also ensures that all the examples are complete from initial planning, up to final execution. The narrative examples actually have more information than what is required for validation, however, this is useful to ensure that the examples can be used for all types of models as only the necessary information can then be used for validation.

The end result of the social engineering attack examples chapter was ten proposed examples. Four of the examples utilised bidirectional communication, three of the scenarios utilised unidirectional communication and three of the scenarios utilised indirect communication. The author ensured that there are several examples from all of the types of communication so that the full spectrum of social engineering attacks are represented in the examples.

The author is of the opinion that the social engineering attack examples are extremely valuable to the field of social engineering. The social engineering attack examples were able to very efficiently validate the SEADM, throughout all of the iterations. The validation was efficient as the author could continuously validate the models with exactly the same parameters allowing the author to have comparative results. The validation was performed easily by just replaying the social engineering attack example through the SEADM.

Ultimately, the author feels that the set of social engineering attack examples should always be increased as more known attacks become available. It will also be very useful if organisations that perform social engineering penetration tests, document their social engineering attacks and provides the anonymised templates of the attack as attack examples back to the community.

The following section focuses on the SEADM. The section discusses the SEADM as a whole and not each iteration individually.

9.7 SOCIAL ENGINEERING ATTACK DETECTION MODEL

The SEADM went through several iterations from when it was first developed in 2010. The first iteration of the SEADM was developed when the author was still very new to the field of social engineering. The first iteration was developed specifically to combat social engineering in a call centre environment. At the time, the author was working within the field of fraud prevention in call centres and realised that call centre agents very often fall prey to social engineering attacks. The model was developed in an effort to assist call centre agents to better screen calls and for the agents to become more aware of the threat of social engineering attacks.

Upon further research into the field of social engineering, specifically when the social engineering definitions were developed, the author learned that social engineering can occur through three means of communication types. This discovery prompted the author to expand on the first iteration of the SEADM by improving the model to cater for all three forms of communication. The second iteration of the SEADM was a significant improvement of the first iteration as it had much better question formulation. The second iteration also focused on equipping the user with the knowledge on how to deal with any type of request and subsequently the user could learn how to detect social engineering attacks without constant need to refer to the model. The idea of states was already included in the second iteration as the states for the request, receiver, requester and third party was already proposed, by means of colour coding. The implementation of the states in the second iteration was poorly implemented as questions related to the request state was proposed in two different parts of the model. This created confusion when one needs to expand on this model as to where questions should be added, hence a third iteration of the SEADM was compiled.

The third and final iteration of the SEADM addressed this problem by having much clearer states. The states were also further expanded upon and the author moved from having four states to having seven states. The states were also explained in much more details as to what each state entails and what questions should be asked in each state. During the improvement of the SEADM it was also noted that the second iteration of the SEADM had an infinite loop at the start of the model and this was addressed by splitting the first part of the model into two separate states and eliminating the infinite loop. The finite state machine of the SEADM was also proven to be deterministic and correct for all possible input alphabets.

The author is of the opinion that the final iteration of the SEADM is the most comprehensive version of the SEADM. The final iteration of the SEADM is also easily extensible since each of the states are clearly described, in this final iteration, as to what their purpose are. In order to extend the SEADM, other researchers may merely add the additional questions at the respective state and the SEADM should still be both deterministic and correct for all input alphabets.

The author performed elaborate validation on all three iterations of the SEADM. Each of the validations are only briefly touched upon in their respective chapters, however, each of the models were tested with the social engineering attack examples that were available at that time. During the development of the first iteration, the author did not yet have the fully fledged social engineering attack examples and only examples from literature was utilised. The second and third iteration of the SEADM was tested using the social engineering attack examples. It was found that both of the iterations was able to correctly detect social engineering attacks, based on the examples utilised.

The author was also part of another research project where the final iteration of the SEADM was implemented as both a web questionnaire and a mobile application. During these studies, the SEADM application was given to participants and the participants were required to determine whether the provided examples are social engineering attacks or not. It should be noted that the research was performed as part of an industry project and the ethical requirements allowed participants to be involved during the process. It was identified that the SEADM greatly assisted the participants to detect social engineering attacks and ultimately, the security awareness of all participants with regards to social engineering had improved [127, 125].

Personally, the author would have preferred to apply the SEADM in a real-world scenario, however, due to the ethical constraints imposed by the University of Pretoria, this was not an option for this thesis. The author attempted to the best of his ability to validate the SEADM by means of example and has even collaborated with other research projects in an effort to further validate the model. The author is extremely happy with the SEADM and is of the opinion that it will eventually be implemented in large organisations in an effort to prevent social engineering attacks. At minimum, the SEADM can always be used as a training tool to educate users on how to detect social engineering attacks. The author did not develop the required training material, however a combination of the SEADM and the social engineering attack examples can easily be converted into training material. The following section concludes this chapter.

9.8 CONCLUSION

This chapter focused on revisiting all of the main contribution topics addressed within this thesis and provided a critical evaluation on each of the contributions. This chapter was divided into sections that discussed each of the main contributions individually. Table 9.1 provides a summary of the contributions and a critical evaluation of each in terms of its advantages and disadvantages.

Table 9.1: Summary of Critical Evaluations on Contributions

Contribution	Advantages	Disadvantages
Standardised definitions for social engineering	The utilisation of standardised definitions by researchers will allow for comparable research.	Researchers may disagree with the proposed definitions.
Ethical concerns regarding social engineering research	Prior to this thesis, there were no ethical guidelines for social engineering research. This thesis proposes guidelines to be followed in future.	The proposed guidelines could be misused to purport that certain ethical perspectives are more favourable to social engineering than others. This is not necessarily a disadvantage, if it constitutes more research into the field.
Social engineering attack framework	This framework provides a guideline on all of the elements that a social engineering attack contains. Furthermore, this framework also provides a consolidated way of depicting all social engineering attacks in a structured manner. This allows for extrapolation of social engineering examples from social engineering attacks in literature.	This framework may be misused by social engineers to improve their existing social engineering attacks.

continued on the next page...

Summary of Critical Evaluations on Contributions		
Contribution	**Advantages**	**Disadvantages**
Social engineering attack examples	The social engineering attack examples provide researchers with detailed examples, containing all of the elements of a social engineering attack with a full narrative on how the attack was performed. The social engineering attack examples also provide researchers with a means to evaluate and validate models within the field of social engineering without the need of ever performing a social engineering attack.	The author is not aware of any disadvantages that the development of social engineering attack examples may entail.
Social Engineering Attack Detection Model	The SEADM can be used to detect social engineering attacks. Whilst individuals are utilising the SEADM, they are also educating themselves to become more vigilant against social engineering attacks and in the future they will be able to detect attacks without the need of referring back to the SEADM. In the event of the SEADM being implemented in an organisation, it will also promote security awareness within the organisation as the question-driven nature of the SEADM causes individual to be much more inquisitive regarding computer security as a whole.	The author is aware that having the SEADM in the public domain may lead attackers to further improve their social engineering attacks. Since the final iteration is a fully extensible model, and the additional questions that can be added does not need to be published, each organisation can customise their question sets to thwart potential attackers.

The author is of the opinion that all of the contributions are of great value to the field of social engineering. The standardised definitions for social engineering is something that can surely be utilised in the field. The author feels that if a large portion of the researchers within the field utilise the standardised definitions, it will allow research within the field to be much more comparable. The research into ethics with regards to social engineering was the first work to examine how different ethical perspectives have an impact within the field of social engineering research. Based on the research performed, the author identified that the different ethical perspectives each have their own interpretation on what is regarded as ethically correct whilst social engineering research is performed. The research contribution allows other researchers to consider the viewpoints of the different ethical perspectives alongside their application for ethical clearance whilst performing social engineering research.

The SEAF and the collection of social engineering attack examples that were extrapolated using the SEAF are novel to the field and it allows for easy validation of any models within the field of social engineering. The social engineering attack examples furthermore provide instances of all three different communication methods that can be utilised for social engineering. This research contribution allows other researchers to validate existing and future models without the need of ever involving participants.

Lastly, the main contribution of this thesis, the SEADM is available to be used as a means to combat social engineering attacks. The SEADM can be used by industry to either implement in their own environment or to be used as a social engineering awareness training tool. The SEADM can also be expanded upon by other researchers and it could also be used as a comparative measure for future social engineering attack detection models.

The next and final chapter, Chapter 10, concludes this thesis by revisiting the research problem and discusses how the research problem was addressed. Chapter 10 also contains references to all the articles and conference papers that orginated from the research process of thesis and a discussion on potential future work is provided.

CHAPTER 10 CONCLUSION

10.1 INTRODUCTION

The focus of this research study was on the development of a Social Engineering Attack Detection Model (SEADM). The proposal of this model required the author to investigate several other fields within social engineering in order to support the development of the model. This thesis started off by describing the problem statement, as well as the research goals and objectives in Chapter 1. Chapter 2 introduced the reader to the existing literature on social engineering, and also proposed consolidated definitions for social engineering which was used throughout this thesis. Chapter 3 introduced the reader to the first iteration of the SEADM. The first iteration was proposed very early on in this thesis in order to provides a baseline for the research objectives of this thesis. Chapter 3 also provided the rationale for further research into all the additional components of social engineering. Chapter 4 discussed the ethical considerations that should be taken into account whilst performing social engineering research. Chapters 5 & 6 provided the reader with the proposed social engineering attack framework (SEAF) and the social engineering attack examples extrapolated from literature using the SEAF. Chapter 7 proposed the second iteration of the SEADM which focused on improving the SEADM by adding the different communication types and the initial idea of states within the model. Chapter 8 proposed the third and final iteration of the SEADM. The final iteration of the SEADM updated the SEADM by converting the model into a finite state machine where each of the states have a unique set of questions linked to it. Chapter 9 critically evaluated all of the research contributions and discussed the value and possible improvements that can be made on each of these contributions.

This last chapter concludes the thesis by revisiting the problem statement and evaluating the extent to which the stated objectives have been accomplished in Section 10.2. Section 10.3 presents and briefly discusses the main contributions of this research study. Section 10.4 provides the reader with a list of

the publications that originated from this thesis while Section 10.5 provides suggestions for potential future research.

10.2 REVISITING THE PROBLEM STATEMENT AND RESEARCH GOALS

This thesis started with a primary problem that *there is currently no formalised method or model for individuals to utilise and especially for educating themselves to be more vigilant against social engineering attacks*. This is an overly complicated problem and the problem was therefore broken up into smaller problems.

The following sub-problems were identified, which needed to be addressed prior to the author continuing to address the overarching problem.

- There is currently no standardised definition for the prominent terms within the field of social engineering, namely social engineer, social engineering, social engineered and social engineering attack.

- There is currently little or no research studies available on the ethical considerations that should be taken into account whilst performing either social engineering attacks or social engineering research.

- There is currently no model that depicts all the phases and steps of a social engineering attack, according to a standardised definition of social engineering attack, onto which already performed or future planned social engineering attacks can be mapped.

- There are limited examples documented of social engineering attacks and most of the documented examples do not include all of the information required to populate all of the phases and steps within the social engineering attack framework.

The thesis first focused on addressing these four sub-problems before the focus was shifted back to the overarching problem concerning the current lack of a formalised method or model for social

engineering attack detection. In order to solve all of these problems, the author identified research goals that this research aimed to achieve.

Since the author identified that there is a primary problem, with sub-problems, the decision was made to also have a primary research goal which focuses on the SEADM and a secondary research goal that focuses on the additional research that can benefit the field of social engineering.

The primary goal of the study was to develop and propose the SEADM. The author had several research objectives in mind that is directly associated with the primary goal:

- The SEADM should be able to cater for various types of social engineering attacks and have no specific focus on a singular type of attack.
- The SEADM should provide a step by step process of identifying social engineering attacks.
- The use of the SEADM must facilitate education and training for users to be more vigilant against social engineering attacks. that will also educate and train users to be more vigilant against such attacks after each request is put through the model.
- The SEADM should be simplistic so that it can be used by any individual without requiring specific training on the model itself.

The secondary goal of this study was to substantially add to and grow the field of social engineering. In order to achieve this goal, the following objectives were pursued:

- The field of social engineering must be enhanced by exploring and proposing standardised definitions.
- Ethical considerations for research in the social engineering domain must be investigated.
- Social engineering attack frameworks must to be investigated.
- Social engineering attack examples must be explored.

All of these topics within the field of social engineering are currently not fully explored and needed to be expanded upon in order to perform the primary goal of this study.

The following section provides the reader with an overview of the contribution of this study within the field of social engineering and how these contributions addressed the problem statement and research goals.

10.3 MAIN CONTRIBUTIONS

This study aimed to make a significant contribution into the field of social engineering. The initial focus of this study was to only make a contribution with regards to the detection of social engineering attacks and how individuals can be more vigilant against social engineering attacks. It was, however, discovered early in the process that this thesis will need to explore a much larger set of topics within the field of social engineering. Therefore, this study addressed formal definitions, ethical considerations, attack frameworks, attack examples and attack detection models within the field of social engineering. The contributions of this study and how they addressed the problem statements are as follows:

1. **Standardised definitions for the terms social engineer, social engineering, social engineered and social engineering attack have been proposed.** (Chapter 2)

    - This contribution addressed the problem that *there is currently no standardised definition for the prominent terms within the field of social engineering, namely social engineer, social engineering, social engineered and social engineering attack.*

    - This contribution thus addressed the secondary research goal of developing the field of social engineering by proposing standardised definitions for terms within the field.

2. **The ethical concerns regarding social engineering from the ethical perspectives of Deontology, Utilitarianism and Virtue Ethics have been defined.** (Chapter 4)

    - This contribution addressed the problem that *there is currently little or no research studies available on the ethical considerations that should be taken into account whilst performing*

either social engineering attacks or social engineering research.

- This contribution thus addressed the secondary research goal of developing the field of social engineering by providing the impact of the three mainstream ethical perspectives on the field of social engineering.

3. **The SEAF which caters for all types of social engineering attacks as encompassed by the standardised definition of social engineering attack has been proposed.** (Chapter 5)

   - This contribution addressed the problem that *there is currently no model that depicts all the phases and steps of a social engineering attack, according to a standardised definition of social engineering attack, onto which already performed or future planned social engineering attacks can be mapped.*

   - This contribution thus addressed the secondary research goal of developing the field of social engineering by proposing a social engineering attack framework that can be utilised to plan and map social engineering attacks.

4. **Ten social engineering attack examples, applied to the social engineering attack framework, which can be used to test and verify the completeness and performance of models within the field of social engineering without requiring to perform the social engineering attacks have been proposed.** (Chapter 6)

   - This contribution addressed the problem that *there are limited examples documented of social engineering attacks and most of the documented examples do not include all of the information required to populate all of the phases and steps within the social engineering attack framework.*

   - This contribution thus addressed the secondary research goal of developing the field of social engineering by documenting a collection of social engineering attack examples that can be utilised to validate models within the field of social engineering.

5. **The SEADM, which can be utilised as a tool by individuals to educate themselves to be more vigilant against social engineering attacks, has been proposed.** (Chapters 3, 7 & 8)

    - This contribution addressed the primary problem that *there is currently no formalised method or model for individuals to utilise and especially for educating themselves to be more vigilant against social engineering attacks.*

    - This contribution also addressed the primary research goal of developing a social engineering attack detection model that adhered to the following research objectives:

        – The first iteration of the SEADM was a model that can be used as a step by step process that will also educate and train users to be more vigilant against social engineering attacks after each request is put through the model.

        – The second iteration of the SEADM, improved on the first one, by being able to cater for all possible types of social engineering attacks and have no specific focus on a singular type of attack.

        – The third and final iteration of the SEADM, improving on the second one, was developed to be simplistic enough so that it can be used by any individual without requiring specific training on the model itself.

        – Care was taken by the author to ensure that none of the qualities were lost throughout each iteration of the SEADM.

This thesis covered a large amount of topics within the field of social engineering. The author ensured that the research became available to other researchers as soon as each of the topics were addressed. Due to this, the author published several articles and conference papers regarding this thesis. The following section provides the reader with a list of these related publications.

10.4 PUBLICATIONS

The author has published a significant amount of work which forms part of this thesis. The decision was made not to include each of the articles as an appendix, however, the full reference is provided in the following list:

- Journals
  - F. Mouton, M. M. Malan, K. K. Kimppa, and H. Venter, "Necessity for ethics in social engineering research," *Computers & Security*, vol. 55, pp. 114 – 127, September 2015. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0167404815001224`
  - F. Mouton, L. Leenen, and H. Venter, "Social engineering attack examples, templates and scenarios," *Computers & Security*, vol. 59, pp. 186 – 209, 2016. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0167404816300268`
  - F. Mouton, A. Nottingham, L. Leenen, and H. Venter, "Finite state machine for the social engineering attack detection model: SEADM," *SAIEE Africa Research Journal*, vol. 109, pp. 133 – 147, 06 2018. [Online]. Available: `http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1991-16962018000200004&nrm=iso`
- Conferences
  - F. Mouton, M. Malan, and H. Venter, "Development of cognitive functioning psychological measures for the seadm," in *Human Aspects of Information Security & Assurance*, Crete, Greece, June 2012, pp. 40–51
  - F. Mouton, M. M. Malan, and H. S. Venter, "Social engineering from a normative ethics perspective," in *Information Security for South Africa*, Johannesburg, South Africa, August 2013, pp. 1–8

- F. Mouton, L. Leenen, M. M. Malan, and H. Venter, "Towards an ontological model defining the social engineering domain," in *ICT and Society*, ser. IFIP Advances in Information and Communication Technology, K. Kimppa, D. Whitehouse, T. Kuusela, and J. Phahlamohlaka, Eds. Springer Berlin Heidelberg, 2014, vol. 431, pp. 266–279. [Online]. Available: `http://dx.doi.org/10.1007/978-3-662-44208-1_22`

- F. Mouton, M. M. Malan, L. Leenen, and H. Venter, "Social engineering attack framework," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2014, pp. 1–9

- F. Mouton, L. Leenen, and H. S. Venter, "Social engineering attack detection model: Seadmv2," in *International Conference on Cyberworlds (CW)*, Visby, Sweden, October 2015, pp. 216–223

- F. Mouton, A. Nottingham, L. Leenen, and H. Venter, "Underlying finite state machine for the social engineering attack detection model," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2017, pp. 1–8

- F. Mouton, M. Teixeira, and T. Meyer, "Benchmarking a mobile implementation of the social engineering prevention training tool," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2017, pp. 1–11

- F. Mouton, M. Pepper, and T. Meyer, "A social engineering prevention training tool: Methodology and design for validating the seadm," in *Human Aspects of Information Security & Assurance*, Dundee, Scotland, August 2018, pp. 12–27

Most of the articles in this list has in some form been discussed already throughout this thesis. It is only the last two articles which does not form part of this thesis as the research performed in those articles were bound to a different set of ethical constraints. The author plans to continue publishing work within the field of social engineering and to return to academia in an attempt to make future researchers aware of the very interesting field of social engineering.

The last section discusses potential future research avenues which can be pursued.

10.5 FUTURE WORK

Computer security, and especially the field of social engineering, is still a growing discipline. The research conducted in this thesis contributes both to computer security and social engineering by exploring and proposing standardised definitions, ethical considerations, attack frameworks, attack examples and attack detection models within the field of social engineering. The main focus of this research study was the SEADM and the author is of the opinion that there is significant room for improvement in all of the other categories.

Whilst compiling the standardised definition, the author also compiled an ontological model for social engineering attacks. The author is of the opinion that this model is still in its infancy and can definitely be expanded upon by other researchers. Additional work is required to fully develop the ontological model. This includes the expansion of classes, as well as the relationships between classes. There is definitely room for improvement of the ontological model if the model is examined by an expert within the field of ontologies.

The SEAF was only used to develop social engineering attack examples and it was never used to perform a social engineering attack as part of a penetration test. The author is of the opinion that the SEAF will greatly benefit by being deployed as a penetration testing tool and any feedback on the usage of the tool will be of valuable research.

There is also room to expand upon the social engineering attack examples. This thesis has only proposed ten attack examples. The author is of the opinion that there is a significantly larger set of examples that can still be developed.

Lastly, with regards to the SEADM, the author is of the opinion that the SEADM has matured to a level where it should now be utilised in the industry. The SEADM has already been developed to be both a web application, as well as an Android mobile application. The application should now be deployed at organisations where key personnel within the organisations utilise the application on a daily basis. The research on exactly how effective the SEADM is within a real-world scenario is something that can definitely be utilised to update and improve the current iteration of the SEADM.

Finally, the author is of the opinion that as technology is becoming more and more accessible to the general public, more social engineering incidents will become prevalent. Almost every single individual has a mobile phone these days, everyone is part of the connected world. The attack platform for social engineers is so much bigger in current times as technology has become so easily accessible. The author is of the opinion that as everyone, even your non-tech savvy users, is part of this connected world and hence the success rate of social engineering attacks is bound to increase. The human element of computer security is of vital importance and constitutes further investigation in protecting individuals.

REFERENCES

[1] J. Quann and P. Belford, "The hack attack - increasing computer system awareness of vulnerability threats," in *3rd Applying Technology to Systems; Aerospace Computer Security Conference*. United States: American Institute of Aeronautics and Astronautics, December 1987, pp. 155–157.

[2] H. Kluepfel, "Foiling the wiley hacker: more than analysis and containment," in *Security Technology, 1989. Proceedings. 1989 International Carnahan Conference on*, 1989, pp. 15–21.

[3] H. Kluepfel, "In search of the cuckoo's nest [computer security]," in *Security Technology, 1991. Proceedings. 25th Annual 1991 IEEE International Carnahan Conference on*, 1991, pp. 181–191.

[4] K. D. Mitnick and W. L. Simon, *The Art of Deception: Controlling the Human Element of Security*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[5] K. D. Mitnick and W. L. Simon, *The Art of Intrusion: The Real Stories Behind the Exploits of Hackers, Intruders & Deceivers*. New York, NY, USA: John Wiley & Sons, Inc., 2005.

[6] C. Hadnagy, *Social Engineering: The Art of Human Hacking*. Wiley Publishing, Inc., 2010.

[7] E. Edelson, "The 419 scam: information warfare on the spam front and a proposal for local filtering," *Computers & Security*, vol. 22, no. 5, pp. 392 – 401, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404803005054

[8] C. Tive, *419 Scam: Exploits of the Nigerian Con Man*. iUniverse, 2006.

[9] F. Mouton, L. Leenen, M. M. Malan, and H. Venter, "Towards an ontological model defining the social engineering domain," in *ICT and Society*, ser. IFIP Advances in Information and Communication Technology, K. Kimppa, D. Whitehouse, T. Kuusela, and J. Phahlamohlaka, Eds. Springer Berlin Heidelberg, 2014, vol. 431, pp. 266–279. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-44208-1_22

[10] M. Workman, "Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 4, pp. 662–674, 2008. [Online]. Available: http://dx.doi.org/10.1002/asi.20779

[11] M. Workman, "A test of interventions for security threats from social engineering," *Information Management & Computer Security*, vol. 16, no. 5, pp. 463–483, 2008.

[12] C. Colwill, "Human factors in information security: The insider threat ï¿½ who can you trust these days?" *Information Security Technical Report*, vol. 14, no. 4, pp. 186 – 196, 2009, human Factors in Information Security. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1363412710000051

[13] The New York Times, "Society topics of the week," *The New York Times*, pp. 3–3, January 1887. [Online]. Available: https://www.nytimes.com/1887/01/02/archives/society-topics-of-the-week.html

[14] The New York Times, "New profession appears; promoters of "social engineering" find a fruitful field," *The New York Times*, p. 8, October 1899. [Online]. Available: https://www.nytimes.com/1899/10/15/archives/new-profession-appears-promoters-of-social-engineering-find-a.html

[15] B. Carlos, "Social engineering, 1899-1999: An odyssey through the new york times," *American Studies in Scandinavia*, vol. 37, no. 1, pp. 69–94, January 2005.

[16] M. Wines, "Bush paints clinton as 'social engineer'," *The New York Times*, p. 21,

September 1992. [Online]. Available: https://www.nytimes.com/1992/09/18/us/the-1992-campaign-the-republicans-bush-paints-clinton-as-social-engineer.html

[17] E. Goldstein, *The Best of 2600, Collector's Edition: A Hacker Odyssey*. Indianapolis, IN: Wiley Publishing, Inc., 2009.

[18] Voyager, "Janitor privileges," *2600: The Hacker Quarterly*, vol. 11, no. 4, pp. 36–36, Winter 1994.

[19] I. S. Winkler and B. Dealy, "Information security technology?...don't rely on it: A case study in social engineering," in *Proceedings of the 5th Conference on USENIX UNIX Security Symposium - Volume 5*, ser. SSYM'95. Berkeley, CA, USA: USENIX Association, 1995, pp. 1–1. [Online]. Available: http://dl.acm.org/citation.cfm?id=1267591.1267592

[20] T. Thornburgh, "Social engineering: the "dark art"," in *Proceedings of the 1st annual conference on Information security curriculum development*, ser. InfoSecCD '04. New York, NY, USA: ACM, 2004, pp. 133–135. [Online]. Available: http://doi.acm.org/10.1145/1059524.1059554

[21] M. Nohlberg, "Securing information assets: Understanding, measuring and protecting against social engineering attacks," Ph.D. dissertation, Stockholm University, 2008.

[22] S. Abraham and I. Chengalur-Smith, "An overview of social engineering malware: Trends, tactics, and implications," *Technology in Society*, vol. 32, no. 3, pp. 183 – 196, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0160791X10000497

[23] M. Erbschloe, *Trojans, worms, and spyware: a computer security professional's guide to malicious code*, Elsevier, Ed. Butterworth-Heinemann, 2004.

[24] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The socialbot network: when bots socialize for fame and money," in *Proceedings of the 27th Annual Computer Security Applications Conference*, ser. ACSAC '11. New York, NY, USA: ACM, 2011, pp. 93–102. [Online]. Available: http://doi.acm.org/10.1145/2076732.2076746

[25] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "Design and analysis of a social botnet," *Computer Networks*, vol. 57, no. 2, pp. 556 – 578, 2013, botnet Activity: Analysis, Detection and Shutdown. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128612002150

[26] D. Kvedar, M. Nettis, and S. P. Fulton, "The use of formal social engineering techniques to identify weaknesses during a computer vulnerability competition," *Journal of Computing Sciences in Colleges*, vol. 26, no. 2, pp. 80–87, Dec. 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1858583.1858595

[27] M. McDowell, "Cyber security tip st04-0141, avoiding social engineering and phishing attacks," United States Computer Emergency Readiness Team, Tech. Rep., February 2013. [Online]. Available: http://www.us-cert.gov/ncas/tips/st04-014

[28] J. A. A. Cruz, "Social engineering and awareness training," Walsh College, Tech. Rep., 2010.

[29] A. M. Culpepper, "Effectiveness of using red teams to identify maritime security vulnerabilities to terrorist attack," Master's thesis, Naval Postgraduate School, Monterey, California, September 2004.

[30] D. Mills, "Analysis of a social engineering threat to information security exacerbated by vulnerabilities exposed through the inherent nature of social networking websites," in *2009 Information Security Curriculum Development Conference*, ser. InfoSecCD '09. New York, NY, USA: ACM, 2009, pp. 139–141. [Online]. Available: http://doi.acm.org/10.1145/1940976.1941003

[31] Q. Doctor, E. Dulaney, and T. Skandier, *CompTIA A+ Complete Study Guide*, W. Publishing, Ed. Indianappolis, Indiana: Wiley Publishing, 2007.

[32] J. T. Hamill, R. F. Deckro, and J. M. K. Jr., "Evaluating information assurance strategies," *Decision Support Systems*, vol. 39, no. 3, pp. 463 – 484, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923604000284

[33] Joint Chiefs of Staff, "Information assurance: Legal, regulatory, policy and organizational

legal, regulatory, policy and organizational considerations," Department of Defense, Pentagon, Washington, Tech. Rep. Fourth Edition, August 1999.

[34] J. T. Hamill, "Modeling information assurance: A value focused thinking approach," Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, March 2000.

[35] M. Braverman, "Behavioural modelling of social engineering-based malicious software," in *Virus Bulletin Conf*, 2006.

[36] R.-M. Åhlfeldt, P. Backlund, B. Wangler, and E. Söderström, "Security issues in health care process integration? a research-in-progress report." in *EMOI-INTEROP*, 2005, pp. 1–4.

[37] S. Granger. (2001, December) Social engineering fundamentals, part i: Hacker tactics. Symantec. [Online]. Available: http://www.symantec.com/connect/articles/social-engineering-fundamentals-part-i-hacker-tactics

[38] A. Schoeman, B. Irwin, and J. Richter, "Social recruiting: a next generation social engineering attack," in *Uses in Warfare and the Safeguarding of Peace*, 2012.

[39] J. Espinhara and U. Albuquerque, "Using online activity as digital fingerprints to create a better spear phisher," Trustwave SpiderLabs, Tech. Rep., 2013.

[40] H. Nemati, *Pervasive Information Security and Privacy Developments: Trends and Advancements*, 1st ed., H. Nemati, Ed. Information Science Reference, July 2010.

[41] S. C. McQuade, III, *Understanding and managing cybercrime*, Allyn and Bacon, Eds. Boston, MA: Prentice Hall, 2006.

[42] M. Spinapolice, "Mitigating the risk of social engineering attacks," Master's thesis, Rochester Institute of Technology B. Thomas Golisano College, 2011.

[43] J. J. Lenkart, "The vulnerability of social networking media and the insider threat new eyes for bad guys," Master's thesis, Naval Postgraduate School, Monterey, California, 2011. [Online].

Available: http://calhoun.nps.edu/public/handle/10945/5562

[44] M. Bezuidenhout, F. Mouton, and H. Venter, "Social engineering attack detection model: Seadm," in *Information Security for South Africa*, Johannesburg, South Africa, August 2010, pp. 1–8.

[45] F. Mouton, M. Malan, and H. Venter, "Development of cognitive functioning psychological measures for the seadm," in *Human Aspects of Information Security & Assurance*, Crete, Greece, June 2012, pp. 40–51.

[46] F. Mouton, M. M. Malan, and H. S. Venter, "Social engineering from a normative ethics perspective," in *Information Security for South Africa*, Johannesburg, South Africa, August 2013, pp. 1–8.

[47] A. Kingsley Ezechi, "Detecting and combating malware," Master's thesis, University of Debrecen, Hungary, June 2011. [Online]. Available: http://hdl.handle.net/2437/105305

[48] D. Harley, "Re-floating the titanic: Dealing with social engineering attacks," in *European Institute for Computer Antivirus Research*, 1998, pp. 4–29.

[49] L. Laribee, "Development of methodical social engineering taxonomy project," Master's thesis, Naval Postgraduate School, Monterey, California, June 2006.

[50] K. Ivaturi and L. Janczewski, "A taxonomy for social engineering attacks," in *International Conference on Information Resources Management*, G. Grant, Ed. Centre for Information Technology, Organizations, and People, June 2011, pp. 1–12.

[51] F. Mohd Foozy, R. Ahmad, M. Abdollah, R. Yusof, and M. Mas'ud, "Generic taxonomy of social engineering attack," in *Malaysian Technical Universities International Conference on Engineering & Technology*, Batu Pahat, Johor, November 2011, pp. 1–7.

[52] P. Tetri and J. Vuorinen, "Dissecting social engineering," *Behaviour & Information Technology*, vol. 32, no. 10, pp. 1014–1023, 2013.

[53] R. Van Rees, "Clarity in the usage of the terms ontology, taxonomy and classification," *CIB Report*, vol. 284, no. 432, pp. 1–8, 2003.

[54] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition - Special issue: Current issues in knowledge modeling*, vol. 5, no. 2, pp. 199–220, June 1993. [Online]. Available: http://dx.doi.org/10.1006/knac.1993.1008

[55] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," Stanford Knowledge Systems Laboratory, Technical Report KSL-01-05, March 2001.

[56] J. W. Scheeres, "Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks," Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, March 2008.

[57] J. Debrosse and D. Harley, "Malice through the looking glass: behaviour analysis for the next decade," in *Proceedings of the 19th Virus Bulletin International Conference*, September 2009.

[58] G. L. Orgill, G. W. Romney, M. G. Bailey, and P. M. Orgill, "The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems," in *Proceedings of the 5th Conference on Information Technology Education*, ser. CITC5 '04. New York, NY, USA: ACM, 2004, pp. 177–181. [Online]. Available: http://doi.acm.org/10.1145/1029533.1029577

[59] D. Gragg, "A multi-level defense against social engineering," SANS Institute InfoSec Reading Room, Tech. Rep., December 2002.

[60] A. N. Chantler and R. Broadhurst, "Social engineering and crime prevention in cyberspace," Queensland University of Technology, Tech. Rep., June 2006. [Online]. Available: http://eprints.qut.edu.au/7526/1/7526.pdf

[61] X. Dong, J. A. Clark, and J. L. Jacob, "User behaviour based phishing websites detection," in *2008 International Multiconference on Computer Science and Information Technology*, Oct 2008, pp. 783–790.

[62] N. Braisby and A. Gellatly, *Cognitive Psychology*, 2nd ed. Oxford University Press, March 2012.

[63] R. J. Sternberg, *Cognitive Psychology*, 4th ed. Wadsworth, 2005.

[64] G. Bansal, F. M. Zahedi, and D. Gefen, "The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online," *Decision Support Systems*, vol. 49, no. 2, pp. 138 – 150, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923610000230

[65] M. T. Siponen, "A conceptual foundation for organizational information security awareness," *Information Management & Computer Security*, vol. 8, no. 1, pp. 31–41, 2000. [Online]. Available: https://doi.org/10.1108/09685220010371394

[66] F. Mouton, M. M. Malan, K. K. Kimppa, and H. Venter, "Necessity for ethics in social engineering research," *Computers & Security*, vol. 55, pp. 114 – 127, September 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404815001224

[67] C. Hadnagy. (2012, December) One royal pwning. Social-Engineer, Inc. [Online]. Available: http://www.social-engineer.org/social-engineering/one-royal-pwning/

[68] L. N. Gowdy. (2013, May) Normative ethics. [Online]. Available: http://www.ethicsmorals.com/ethicsnormative.html

[69] G. Harman, "Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error," in *Proceedings of the Aristotelian Society*, vol. 99, JSTOR. Wiley on behalf of The Aristotelian Society, 1999, pp. 315–331.

[70] I. Manners, "The normative ethics of the european union," *International Affairs*, vol. 84, no. 1, pp. 45–60, 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1468-2346.2008.00688.x

[71] A. Stevenson, *Oxford Dictionary of English*, ser. Oxford reference online premium. OUP Oxford, 2010. [Online]. Available: http://books.google.co.za/books?id=TaZZSAAACAAJ

[72] D. Knights and M. O?Leary, "Leadership, ethics and responsibility to the other," *Journal of Business Ethics*, vol. 67, no. 2, pp. 125–137, 2006. [Online]. Available: http://dx.doi.org/10.1007/s10551-006-9008-6

[73] N. Athanassoulis. (2014, July) Virtue ethics. Keele University. United Kingdom. [Online]. Available: http://www.iep.utm.edu/virtue/

[74] W. S. Sahakian and M. L. Sahakian, *Ideas of the great philosophers*. Barnes & Noble Publishing, 1966, no. 218.

[75] H. Simmons. (2013, September) Guide to moral philosophy. British Philosophical Association. [Online]. Available: http://philosophyadvice.net/Guide.pdf

[76] IEEE Board of Directors. (2014, August) Ieee policies. The Institute of Electrical and Electronics Engineers. New York, USA. [Online]. Available: http://www.ieee.org/documents/ieee_policies.pdf

[77] ACM Council. (1992, October) Acm code of ethics and professional conduct. ACM. [Online]. Available: http://www.acm.org/about/code-of-ethics

[78] ComputingCases. (2014, February) Publicity test. ComputingCases.org. [Online]. Available: http://www.computingcases.org/general_tools/teaching_with_cases/ethics_tests/publicity_test.html

[79] BBC. (2014, February) Consequentialism. BBC. [Online]. Available: http://www.bbc.co.uk/ethics/introduction/consequentialism_1.shtml

[80] BBC. (2014, February) Duty-based ethics. BBC. [Online]. Available: http://www.bbc.co.uk/ethics/introduction/duty_1.shtml

[81] L. Alexander and M. Moore, "Deontological ethics," in *The Stanford Encyclopedia of Philosophy*, 1st ed., E. N. Zalta, Ed. Stanford, 2012.

[82] C. Hadnagy. (2014, February) Social engineering penetration testing. Social-Engineer, Inc. [Online]. Available: http://www.social-engineer.com/social-engineer-pentesting/

[83] D. Simpson. (2014, February) @whackheads. [Online]. Available: http://twitter.com/WhackheadS

[84] M. Theunissen. (2013, January) Marietta theunisse biography. Other Worlds Tomorrow. [Online]. Available: http://www.otherworldstomorrow.com/upload/MariettaTheunissenBioJan2013.pdf

[85] C. Hadnagy. (2010, June) Social engineering: Past, present and future. Social-Engineer, Inc. [Online]. Available: http://www.social-engineer.org/episode-010-social-engineering-past-present-and-future/

[86] W. Kearney and H. Kruger, "Considering the influence of human trust in practical social engineering exercises," in *Information Security for South Africa*, Johannesburg, August 2014, pp. 1–6.

[87] F. Mouton, M. M. Malan, L. Leenen, and H. Venter, "Social engineering attack framework," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2014, pp. 1–9.

[88] R. Cialdini, *Influence: The Psychology of Persuasion*, R. Cialdini, Ed. HarperCollins Publishers, 2007.

[89] J. Long, *No tech hacking: A guide to social engineering, dumpster diving, and shoulder surfing*, S. Pinzon, Ed. Syngress, 2011.

[90] Symantec Security Response. (2014, January) Francophoned ? a sophisticated social engineering attack. Symantec. [Online]. Available: http://www.symantec.com/connect/blogs/francophoned-sophisticated-social-engineering-attack

[91] L. Zeltser. (2009, February) Malware infection that began with windshield fliers. Internet Storm Center. [Online]. Available: https://isc.sans.edu/diary/5797

[92] F. Mouton, L. Leenen, and H. Venter, "Social engineering attack examples, templates and scenarios," *Computers & Security*, vol. 59, pp. 186 – 209, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404816300268

[93] L. Janczewski and L. Fu, "Social engineering-based attacks: Model and new zealand perspective," in *Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on*, Oct 2010, pp. 847–853.

[94] T. Dimkov, A. van Cleeff, W. Pieters, and P. Hartel, "Two methodologies for physical penetration testing using social engineering," in *Proceedings of the 26th Annual Computer Security Applications Conference*, ser. ACSAC '10. New York, NY, USA: ACM, 2010, pp. 399–408. [Online]. Available: http://doi.acm.org/10.1145/1920261.1920319

[95] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung, "Fourth-factor authentication: Somebody you know," in *Proceedings of the 13th ACM Conference on Computer and Communications Security*, ser. CCS '06. New York, NY, USA: ACM, 2006, pp. 168–178. [Online]. Available: http://doi.acm.org/10.1145/1180405.1180427

[96] R. G. Brody, W. B. Brizzee, and L. Cano, "Flying under the radar: social engineering," *International Journal of Accounting & Information Management*, vol. 20, no. 4, pp. 335–347, 2012.

[97] S. D. A. Major, "Social engineering: Hacking the wetware!" *Information Security Journal: A Global Perspective*, vol. 18, no. 1, pp. 40–46, 2009.

[98] J. Jetten, M. J. Hornsey, and I. Adarves-Yorno, "When group members admit to being conformist: The role of relative intragroup status in conformity self-reports," *Personality and Social Psychology Bulletin*, vol. 32, no. 2, pp. 162–173, February 2006.

[99] H. A. Simon, *Models of man; social and rational.* Oxford, England: Wiley, 1957.

[100] D. Hill, "Peer group conformity in adolescent smoking and its relationship to affiliation and autonomy needs," *Australian Journal of Psychology*, vol. 23, no. 2, pp. 189–199, 1971. [Online].

Available: http://www.tandfonline.com/doi/abs/10.1080/00049537108254613

[101] H. B. Gerard, R. A. Wilhelmy, and E. S. Conolley, "Conformity and group size." *Journal of Personality and Social Psychology*, vol. 8, no. 1p1, pp. 79–82, January 1968.

[102] A. J. Lott and B. E. Lott, "Group cohesiveness, communication level, and conformity." *The Journal of Abnormal and Social Psychology*, vol. 62, no. 2, pp. 408–412, March 1961.

[103] J. E. Dittes and H. H. Kelley, "Effects of different conditions of acceptance upon conformity to group norms." *The Journal of Abnormal and Social Psychology*, vol. 53, no. 1, pp. 100–107, July 1956.

[104] C. A. Insko, R. H. Smith, M. D. Alicke, J. Wade, and S. Taylor, "Conformity and group size the concern with being right and the concern with being liked," *Personality and Social Psychology Bulletin*, vol. 11, no. 1, pp. 41–50, March 1985.

[105] G. Bader, A. Anjomshoaa, and A. Tjoa, "Privacy aspects of mashup architecture," in *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, Aug 2010, pp. 1141–1146.

[106] L. Tam, M. Glassman, and M. Vandenwauver, "The psychology of password management: a tradeoff between security and convenience," *Behaviour & Information Technology*, vol. 29, no. 3, pp. 233–244, 2010. [Online]. Available: http://dx.doi.org/10.1080/01449290903121386

[107] T. R. Peltier, "Social engineering: Concepts and solutions," *Information Systems Security*, vol. 15, no. 5, pp. 13–21, January 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1201/1086.1065898X/46353.15.4.20060901/95427.3

[108] U. Rao and U. Nayak, "Social engineering," in *The InfoSec Handbook*. Apress, 2014, pp. 307–323. [Online]. Available: http://dx.doi.org/10.1007/978-1-4302-6383-8_15

[109] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, Oct. 2007. [Online]. Available: http://doi.acm.org/10.1145/

1290958.1290968

[110] H. Jahankhani, "The behaviour and perceptions of on-line consumers: Risk, risk perception and trust," *International Journal of Information Science and Management*, vol. 7, no. 1, pp. 79–90, June 2012.

[111] O. Salem, A. Hossain, and M. Kamala, "Awareness program and ai based tool to reduce risk of phishing attacks," in *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, June 2010, pp. 1418–1423.

[112] S. Schrittwieser, P. Frühwirt, P. Kieseberg, M. Leithner, M. Mulazzani, M. Huber, and E. R. Weippl, "Guess who's texting you? evaluating the security of smartphone messaging applications," in *Network and Distributed System Security Symposium*, February 2012, pp. 1–9.

[113] K. Krombholz, H. Hobel, M. Huber, and E. Weippl, "Social engineering attacks on the knowledge worker," in *Proceedings of the 6th International Conference on Security of Information and Networks*, ser. SIN '13. New York, NY, USA: ACM, 2013, pp. 28–35. [Online]. Available: http://doi.acm.org/10.1145/2523514.2523596

[114] H. Dang, "The origins of social engineering," *McAfee Security Journal*, vol. 1, no. 1, pp. 4–9, Fall 2008.

[115] S. Stasiukonis. (2006, June) Social engineering, the usb way. Dark Reading. [Online]. Available: http://tonydye.typepad.com/main/files/HO05-DarkReading.doc

[116] S. Esmail, "eps1.5_br4ve-trave1er.asf," June 2015, mr. Robot: Season 1, Episode 6. [Online]. Available: http://www.usanetwork.com/mrrobot/episode-guide/season-1-episode-6-eps15br4ve-trave1erasf

[117] M. Jodeit and M. Johns, "Usb device drivers: A stepping stone into your kernel," in *Computer Network Defense (EC2ND), 2010 European Conference on*, Oct 2010, pp. 46–52.

[118] G. Brown, T. Howe, M. Ihbe, A. Prakash, and K. Borders, "Social networks and context-aware spam," in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '08. New York, NY, USA: ACM, 2008, pp. 403–412. [Online]. Available: http://doi.acm.org/10.1145/1460563.1460628

[119] D. Irani, M. Balduzzi, D. Balzarotti, E. Kirda, and C. Pu, "Reverse social engineering attacks in online social networks," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, ser. Lecture Notes in Computer Science, T. Holz and H. Bos, Eds. Springer Berlin Heidelberg, 2011, vol. 6739, pp. 55–74. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-22424-9_4

[120] P. Kieseberg, M. Leithner, M. Mulazzani, L. Munroe, S. Schrittwieser, M. Sinha, and E. Weippl, "Qr code security," in *Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia*, ser. MoMM '10. New York, NY, USA: ACM, 2010, pp. 430–435. [Online]. Available: http://doi.acm.org/10.1145/1971519.1971593

[121] N. Gruschka and M. Jensen, "Attack surfaces: A taxonomy for attacks on cloud services," in *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*. Los Alamitos, CA, USA: IEEE Computer Society, July 2010, pp. 276–279.

[122] F. Mouton, L. Leenen, and H. S. Venter, "Social engineering attack detection model: Seadmv2," in *International Conference on Cyberworlds (CW)*, Visby, Sweden, October 2015, pp. 216–223.

[123] R. Bhakta and I. Harris, "Semantic analysis of dialogs to detect social engineering attacks," in *Semantic Computing (ICSC), 2015 IEEE International Conference on*, Feb 2015, pp. 424–427.

[124] F. Mouton, A. Nottingham, L. Leenen, and H. Venter, "Underlying finite state machine for the social engineering attack detection model," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2017, pp. 1–8.

[125] F. Mouton, A. Nottingham, L. Leenen, and H. Venter, "Finite state machine for the social engineering attack detection model: SEADM," *SAIEE Africa Research Journal*, vol. 109,

pp. 133 – 147, 06 2018. [Online]. Available: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1991-16962018000200004&nrm=iso

[126] S. S. Epp, *Discrete Mathematics with Applications*, 4th ed. Brooks/Cole Publishing Co., 2010.

[127] F. Mouton, M. Teixeira, and T. Meyer, "Benchmarking a mobile implementation of the social engineering prevention training tool," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2017, pp. 1–11.

[128] F. Mouton, M. Pepper, and T. Meyer, "A social engineering prevention training tool: Methodology and design for validating the seadm," in *Human Aspects of Information Security & Assurance*, Dundee, Scotland, August 2018, pp. 12–27.