A Social Engineering Prevention Training Tool: Methodology and Design for Validating the SEADM

F. Mouton[1], M. J. Pepper[2] and T. Meyer[3]

[1]Defence Peace Safety & Security, [2,3]Department of Computer Science
[2]Council for Scientific and Industrial Research, [2,3]University of Cape Town,
[3]Pretoria, South Africa, [2,3]Cape Town, South Africa
e-mail: moutonf@gmail.com; mikejpepper@gmail.com; tmeyer@cs.uct.ac.za

Abstract

The information people possess is often of great value and thus, when stored electronically, is typically guarded by complicated security mechanisms. Such mechanisms are frequently upgraded in order to counteract threats that aim to obtain the information being guarded. Accordingly, the "social engineer" seeks to attack and exploit the weakest link in this information security system: the user. The general public is often not aware that they may be subjected to acts of social engineering (SE), and are hence not aware of what to look for and how to react appropriately in such situations. This leaves the unsuspecting public in a vulnerable position with very little assistance at their disposal.

The Social Engineering Prevention Training Tool (SEPTT) project of which we are part sought to address SE vulnerability by developing a tool that can be used in any scenario to determine if the user is being subjected to acts of SE, and to provide guidance as to the correct manner of response to follow in said scenario. The authors previously expanded on the original Social Engineering Attack Detection Model and produced the updated version 2, i.e. SEADMv2. A test methodology to validate the updated model is presented together with a preliminary design for the web application.

Keywords

Social engineering, attack prevention, Mitnick's attack cycle, SEADMv2, social engineering attack detection model, social engineering attack examples, social engineering attack framework.

1 Introduction

Social engineers make use of psychological ploys that compromise the user's emotional state, hence allowing an "exploit" to take place (Bezuidenhout, Mouton, & Venter, 2010; Mouton, Leenen, Malan, & Venter, 2014; Mouton, Malan, et al., 2014). This psychological manipulation can be performed using various techniques through multiple channels and mediums. However, the overall goal is the same. By exploiting psychological vulnerabilities of users, social engineers destabilise users' thinking so as to elicit responses – and hence perform information-gathering – that would not be possible had the user been in a more stable state of mind (Bezuidenhout et al., 2010; Mouton et al., 2012). This ultimately leads to the attacker achieving a predetermined objective, often unbeknownst to the victim. The success of these attacks can often be attributed to individuals not perceiving themselves as potential victims of such attacks

and hence not being aware of the types of techniques used in their execution (Mouton, Leenen, & Venter, 2015). This ignorance may be due to the users' lack of knowledge of the potential gains an attacker can attain from the information they possess.

The 'art' of influencing people to divulge sensitive information is known as social engineering and the process of doing so is known as a social engineering attack. There are various definitions of social engineering and also a number of different models of social engineering attack (Mitnick and Simon, 2002; Culpepper, 2004; Thornburgh, 2004; Åhlfeldt *et al.*, 2005; Hamill, Deckro and Jr., 2005; Nohlberg, 2008; Hadnagy, 2010; Kingsley Ezechi, 2011; Lenkart, 2011; Mouton, Leenen, *et al.*, 2014). The authors considered a number of definitions of social engineering and social engineering attack taxonomies in a previous paper, *Towards an Ontological Model Defining the Social Engineering Domain*, and formulated a definition for both social engineering and social engineering attack (Mouton, Leenen, *et al.*, 2014). In addition, the authors proposed an ontological model for a social engineering attack. It is important to ensure that a standardised definition is used throughout all the work within a single domain. For the purpose of this paper, definition for social engineering used throughout this paper 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" (Mouton, Leenen, *et al.*, 2014).

There has been a significant amount of research performed into defining the field of social engineering and furthermore social engineering prevention, there has not yet been any research into the development of a tool. As far as the authors are aware, there is currently no tool available that can be used to detect social engineering attacks and give users an indication of the action they should take in a given scenario. This naturally leaves people in a vulnerable position, with the only assistance available to them being generic 'tips' on things to look out for. The Social Engineering Prevention Training Tool (SEPTT) project of which we are part aims at addressing this gap by implementing the Social Engineering Attack Detection Model Version 2 (SEADMv2) proposed by (Mouton et al., 2015) as a web application, in order to determine whether it is effective at assisting users to successfully differentiate between harmless requests and genuine SE attacks. SEADMv2 aims to guide its users towards understanding of the appropriate action to take in given scenarios, hence reducing the probability of them falling victim to an SE attack.

2 Background

This section analyses the current frameworks available to model SE attacks, with emphasis on the framework proposed by Mitnick and Simon (2002; 2005). The differing SE attack classifications are also outlined, as they are pivotal in creating SE attack scenarios that accurately depict real-world attacks for the experiment to follow.

2.1 Mitnick's Attack Cycle

In order to combat the vulnerability of the unsuspecting public, the first step is to understand how SE attacks are structured so that each aspect of the attack can be accounted for. Mitnick's attack cycle is pivotal in this regard as it is the most widely accepted SE attack framework, since its phases are consistent across all attack types (Mitnick and Simon, 2005). The cycle breaks an SE attack down into several phases, each of which contains a predetermined goal. These phases are discussed in the following subsections, with reference to alternate models that define similar phases.

2.1.1 Information-gathering

Initially, the social engineer gathers as much information about the target as possible (Mouton, Malan, et al., 2014). This information-gathering can take many forms and aims at acquiring information and resources necessary to successfully perform an attack (Van De Merwe and Mouton, 2017). The quality of information attained plays a vital role in successfully creating a relationship with the target, a stage that is pivotal in the overall success of the attack (Mouton, Malan, et al., 2014). Techniques such as gathering Facebook pictures of the target's friends and identifying the language and tone used between the target and those friends are examples of techniques that could be used in this phase (Abraham and Chengalur-Smith, 2010). Such information would assist in masquerading as one of the target's friends in order to exploit their relationship and attain valuable information from that individual.

2.1.2 Develop rapport and trust

Once sufficient information is gathered about the target, the social engineer attempts to establish a relationship with the target as the target will be more likely to divulge the requested information to the attacker if there is an existing relationship (Mouton, Malan, et al., 2014). Developing this relationship relies on the information gathered in the previous phase, as the approach used is tailored to the information available. For example, social engineers may use insider information to masquerade as someone within an organisation; misrepresent their identity by pretending to be a specific individual; cite individuals known by the target as common connections aid in an individual's credibility; or appear to occupy an authoritative role (Mouton, Malan, et al., 2014). In doing this, the attacker hopes to establish some trust connection with the target (Gao and Kim, 2007), which will make the target more susceptible to exploitation in the next phase.

2.1.3 Exploit trust

Once a relationship has been established, the attacker attempts to exploit this trust to gain information from the target. In Mitnick's attack-cycle model, this is achieved through manipulation of the target's emotional state by preying on the seven psychological vulnerabilities noted by Gragg (2002). They are: strong affect, overloading, reciprocation, deceptive relationship, diffusion of responsibility and moral duty, authority, integrity and consistency (Mitnick and Simon, 2005; Chantler and Broadhurst, 2006; Scheeres, 2008; Workman, 2008). By exploiting these

psychological vulnerabilities, the target's emotional state is altered and she or he becomes more likely to comply with the attacker's requests for information (Mouton, Malan, et al., 2014).

2.1.4 Utilise information

Lastly, Mitnick's model notes the phase in which the information gathered in the previous phase is utilised to achieve the predefined goal (Mitnick and Simon, 2005). Should insufficient information be attained, the model cycles back to phase one. Other models fail to recognise this phase and deem the social engineering attack to be successful once the required information is retrieved from the target.

2.2 Attack Classifications

SE attacks can be classified according to the manner in which the communication takes place during the exploit, and the interaction between attacker and target (Mouton, Malan, et al., 2014). By understanding the different types of attacks, one can generate attack scenarios representative of possible real-life attacks, with a broad enough coverage to account for the differing manners in which these are performed.

According to Mouton, Leenen et al. (2014), SE attacks can be divided into direct and indirect attacks. In this classification, indirect attacks are those where a third-party medium is used to facilitate the communication between attacker and target. In such attacks, communication takes place when a target accesses the third party medium without interaction from the social engineer. Mediums such as USB flash drives and pamphlets are used to exploit the target in some way (Abraham and Chengalur-Smith, 2010).

Direct attacks are those where two or more parties are involved in a direct conversation. Direct attacks are differentiated in this model on whether they are one-sided or two-sided. One-sided attacks are classified as *unidirectional communication* and two-sided as *bidirectional communication*. Bidirectional communication takes place when two or more parties partake in a conversation. . This type of communication can be likened to the communication described by Ivaturi and Janczewski (2011, 2012) and is often performed over interactive media such as email and face-to-face conversations as both parties need to be able to contribute. Unidirectional communication occurs when there is communication between attacker and target without the target being able to converse with the attacker in a back-and-forth manner. Examples of the media used for such communication are emails and one-way text messages. Diagrams depicting these different types of communication can be found in a paper by Mouton *et al.* (2014) entitled, *Towards an Ontological Model Defining the Social Engineering Domain.*

3 Social Engineering Attack Detection Model Version 2 (SEADMv2)

The SEADMv2 (Mouton et al., 2015) is a revision of the model initially proposed by Bezuidenhout et al. (2010). This revised model provides users with a state diagram that can be used to determine: firstly, if they are being subjected to acts of SE; and secondly, the appropriate action they should take. It achieves this by asking the users questions about their current scenario, the answers to which determine their transitions through the model (seen in Figure 1 below). The model eventually reaches a termination state, at which point the user is given one of two instructions: "perform the request" or "defer or refer request". The instruction to "perform the request" indicates to the user that she or he should comply with the requester's demands and perform the relevant action as it is unlikely to be an SE attack. The instruction to "defer or refer request" indicates to the user that she or he may be subjected to an SE attack and should thus *refer* the request to someone better-suited to deal with it, or *defer* the request completely – whichever would be more applicable to the user in a real-life situation.

This version 2 of the SEADM improves upon the Bezuidenhout et al. (2010) first iteration through expanding upon the states proposed, hence increasing the model's coverage and making it more user-friendly. Additionally, the 'state' component in the previous model, which required the user to evaluate his or her emotional state, has been omitted and is now dealt with by a separate psychological measure developed by Mouton et al. (2012).

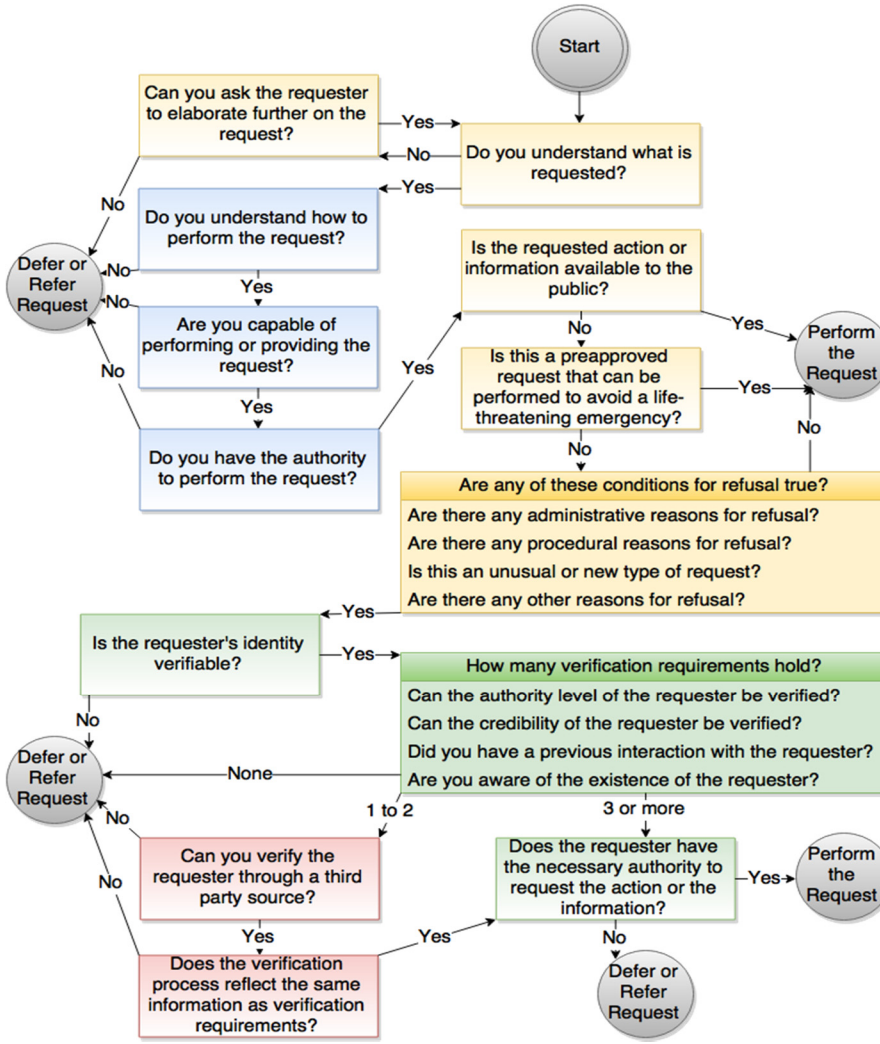

Figure 1: Social Engineering Attack Detection Model version 2 (SEADMv2) (Mouton, Leenen and Venter, 2015)

The information on how to process each of the states is discussed in an article by Mouton, Leenen and Venter (2015) entitled, "Social Engineering Attack Detection Model: SEADMv2". The SEADMv2 has also been further developed into a finite state machine, where the colour coded areas of the SEADMv2 is further reduced to a set of states (Mouton *et al.*, 2017, 2018). This allows the model to be fully extensible and allows one to further ask more questions per state and is thus not limited to the predefined set of questions. The designed web implementation of the SEPTT caters for all the rules of the finite state machine. The finite state machine is depicted in Figure 2.

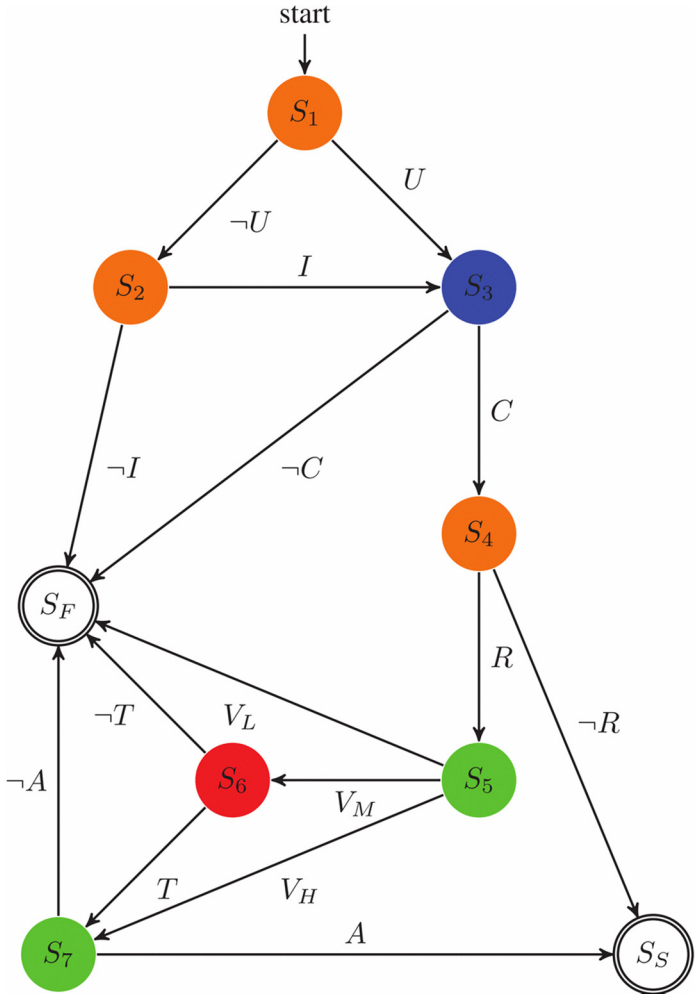

Figure 2: Finite State Machine of the SEADM (Mouton et al., 2017, 2018)

4 Methodology

A two-stage experiment is proposed with an identified 45 subjects. Subjects will be sent a link to the questionnaire and invited to participate in a two stage experiment and provided with instructions on how to go about completing the experiment. The order in which questions were asked in both stages was randomised to avoid any ordering effects on subjects' answers. The two stages of the experiment are discussed below, as well as the data transformation that was performed to transform the results to a usable format for statistical testing.

Proposed Experiment

Stage 1: The first stage will consist of the 10 potential SE attack scenarios mentioned above, each with four possible answers: two "perform the request" options and two "defer or refer the request" options. For example, in order to choose to "defer or refer the request", the subject had to choose a multiple-choice option that did not comply with the requests in the scenario, or that deferred the situation to someone better-equipped to deal with it. These answers will provide a record of how subjects respond to each scenario without assistance from the SEADMv2-model application. This will form the "without model" before-treatment data collection stage, and serve as the control results of the experiment.

Stage 2: Upon completion of stage 1, the subjects will be informed that they must now make use of the SEADMv2 web model to guide their answers to the previous 10 scenarios. To achieve this, the same 10 scenarios will be presented to the subjects in a random order. However now, for each scenario, they would have to use the information in that scenario to progress through the SEADMv2 model by answering "yes" or "no" to the questions it asked. The result of this stage of the experiment will be a record of how subjects react to each scenario when they have the guidance of the SEADMv2 model and constitutes the "With Model" after-treatment data.

Responses to the questionnaire will be limited to one per person to prevent the same person answering it multiple times and skewing the data.

5 Proposed Design and Implementation

The hypotheses that this experiment seeks to test are:

- that user interaction with the SEADMv2 web application will significantly increase the user's ability to recognise and avoid genuine SE attack requests; and
- user interaction with the SEADMv2 web application will significantly increase the user's ability to recognise and reply favourably towards harmless requests.

The efficacy of the model will be assessed through a two-stage experiment, whereby subjects will be given 10 scenarios that are possible social engineering attacks, with four possible options of how to respond to each scenario. Subjects had to choose the option that most accurately depicted how they would react in each scenario, first without the use of the SEADMv2 model (stage 1 of the experiment) and then (in stage 2) with use of the SEADMv2 model.

To perform this experiment, a web application will be created that allows users to traverse the SEADMv2. The research subjects who will use the tool have been identified, and necessary consent and ethical issues are being duly managed.

This section discusses the design considerations and techniques that will be employed to develop the application, as well as the scenarios that were created to assess the tool's efficacy. The questionnaire through which the experiment was conducted is also discussed.

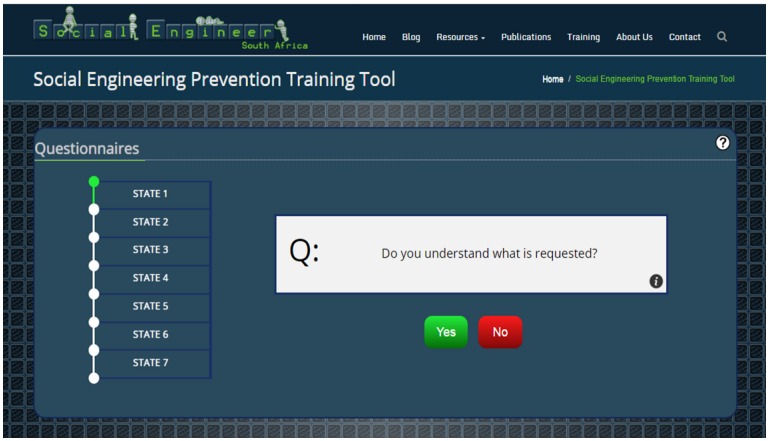

Figure 3: Social Engineering Prevention Training Tool (Web Application)

5.1 Web application

The web application (seen above in Figure 3) consists of a question box that poses a question to the user that is dependent on the user's current state within the SEADMv2 model. Each question aims to assess the user's knowledge of the current situation before the user transitions to the next state in the model. (Eventually, at the final stage, the user has decided on the correct action to take.) Below each question, there are two buttons that allow the users to answer "yes" or "no". There is a progress bar on the left side of the interface indicating to the users their current position in the model, and informational buttons that can be used to aid the users in the event that they seek clarity on some aspect of the current question. A rapid application design (RAD) approach was used to develop this web application, hence ensuring that the resulting application was developed to specification and within time constraints. The web application will be hosted on www.social-engineering.co.za and make use of a MySQL database to store the SEADMv2 model.

5.2 Social engineering scenarios

Ten believable real-life situations were drafted into scenario format. These scenarios focused on two phases of Mitnick's attack-cycle model, namely the *develop rapport and trust* phase and the *exploit trust* phase (Mitnick and Simon, 2002). The scenarios also employed request techniques used within these phases by successful SE attackers. Each request scenario terminated with four possible responses for the user to choose from: two responses that signalled compliance with the request (i.e., responses indicating that the user felt the request was not an SE attack), and two responses that did not comply with the request (i.e., two responses indicating user suspected the request was an SE attack).

The 10 scenarios comprised eight that were characteristic of genuine SE attacks, and two that were characteristic of harmless requests that could be complied with. After completing the experiment, it became clear that a more even split between genuine-attack scenarios and harmless scenarios would have been ideal, as the low number of harmless scenarios affected the credibility of those results. This lack of foresight, as discussed in the "limitations" sub-section below, arose during the planning stages of the experiment when the only consideration was that there should be harmless scenarios, with insufficient consideration of how many there should be in relation to the number of potentially harmful. This led to the less-than-ideal 8/2 split.

In order to ensure that the scenarios are diverse enough to model the different types of real-world attacks, the SE attack classifications developed by Mouton, Leenen, and Venter (2016) is to be used as templates. Of the ten scenarios that were created for the experiment, five depicted unidirectional communication, four depicted bidirectional communication, and one depicted indirect communication. To provide the reader with a sense of the content and structure of the scenarios, five of the ten are briefly outlined in the section below and includes the following types:

- two unidirectional communication scenarios;
- two bidirectional communication scenarios; and
- one indirect communication scenario.

5.2.1 Unidirectional communication scenarios

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 (Mouton, Leenen, et al., 2014).

5.2.2 Scenario 1

Summary**:** While at work you receive an email from a new email address indicating that a new person (the sender) from your company's external accounting firm has started working on the time reports for this quarter and hence she needs you to send your preliminary time report through as soon as possible. The email address that the message comes from has the same domain as previous emails from the accounting firm and the signature of the email is the same as all previous emails from various other employees of the accounting firm. What action do you take?

Notable aspects of scenario: you understand how to perform the request; you are capable of performing the request and have the authority to do so; information requested is sensitive and not publicly available; this is a unique request and not pre-authorised; there are administrative reasons to not perform this request; the requester's identity, authority and credibility are verifiable; you have had no previous interaction with the requester but can verify the requester's intentions.

Possible responses to scenario:

A. since you do not have much work to do, you get working on your preliminary time report immediately and email it to the requester as soon as possible.
B. you reply to the email, asking her a few complementary questions and, based on her answers, either provide her your preliminary time report or refuse to send it to her.
C. you contact your superior to find out whether or not they approve of you sending your preliminary time report to the person requesting it.
D. you refuse to send her your preliminary time report.

Suggested (i.e., secure) action: Perform the request, i.e., choose option A or B.

5.2.3 Scenario 2

Summary**:** Whilst sitting in a lecture at university, your lecturer introduces a guest lecturer from an external organisation. The guest lecturer gives a bit of information about his organisation and hands out a small assignment that will count towards your final grade at the end of the year. The assignment asks for your student number as well as date of birth and last seven digits of your identification document (ID) number. The guest lecturer assures you that the information will only be used for recruitment purposes. What action do you take?

Notable aspects of scenario: You understand how to perform the request; You are capable of performing the request and have the authority to do so; Information requested is not available to the public; This is not a pre-approved request; There are administrative reasons for refusal; The requester's identity is not verifiable.

Possible responses to scenario:

A. you provide all the requested information.
B. you ask the guest lecturer a few complementary questions and based on his answers decide whether to provide the information.
C. you ask the guest lecturer to rather contact your lecturer directly to obtain this information.
D. you do not provide the information and also do not tell the guest lecturer where to get it as you deem it to be sensitive information.

Suggested (i.e., secure) action: Defer or refer request, i.e., choose option C or D.

5.2.4 Bidirectional communication scenarios

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 (Mouton, Leenen, et al., 2014).

5.2.5 Scenario 3

Summary**:** You receive a message on Facebook from a person you do not know who claims he is a marketing agent for the Rocking the Daisies Festival. The message tells you about a competition to win free tickets to the festival. All that is required is that you send through a video explaining how excited you are about the festival and why you think you should win. You verify that there is in fact a competition to win tickets by going onto the Rocking the Daisies Facebook page and seeing the competition advertised as the person explained. The message states further that they would like to assist you with your entry as they receive commission for each entry they provide assistance to. To do this, they ask that you send your video to them directly, along with your full name, date of birth and Facebook login details (email and password), since an entry requires a link to your Facebook account. What action do you take?

Notable aspects of scenario: You understand how to perform the request; you are capable of performing the request and have the authority to do so; information requested is sensitive and not publicly available; this is a new type of request and not pre-authorised; there are administrative reasons for refusal; the requester's identity is not verifiable.

Possible responses to scenario:

A. you record your video in a few days and send him your video along with all the information requested, since he only needs it to enter you into the competition.
B. you record your video in a few days and send him your video along with all the information requested (however you are a bit wary about giving out your Facebook login details and decide to change your Facebook password 24 hours after sending it to him).
C. you record your video in a few days, but decide to rather enter the competition yourself by going to the official festival website and entering the competition there, without sending the person who contacted you on Facebook any of your details.
D. you decide not to enter the competition at all (since the person on Facebook was asking for your Facebook login details for the competition, you conclude that the entire competition must be fake and decide that it is best not to enter).

Suggested (i.e., secure) action: Defer or refer request, i.e., choose option C or D.

5.2.6 Scenario 4

Summary**:** As a university student, you are walking to the turnstile entrance of the computer lab when a person you do not know approaches you. The person looks like a student and asks you to swipe them through the turnstile using your student card as they have forgotten theirs at home. You know that swiping in other students to labs is not allowed, but you can see that the student is stressed and has an assignment to submit within the next 15 minutes. What action do you take?

Notable aspects of scenario: You understand how to perform the request; you are capable of performing the request; you do not have the authority to perform the request.

Possible responses to scenario:

A. you swipe the student in immediately, since you know how stressful it is submitting an assignment at the last minute and you know there is no time to waste.
B. even though the student is stressed and needs to get into the lab as soon as possible, you decide to ask the student a few questions and based on his/her answers make a decision on whether to swipe him/her in or not.
C. you refuse to help the student at all and tell the student he/she should not have waited until the last minute to submit the assignment and he/she should always have their student card on them while on campus.
D. you give the student directions to the access control offices where the student can prove his/her identity and hopefully get access to a computer lab within 15 minutes to submit the assignment.

Suggested (i.e., secure) action: Defer or refer request, i.e., choose option C or D.

5.2.7 Indirect communication scenario

Indirect communication occurs when a third-party medium is used as a form of transporting the communication. Typical third-party media include physical media such as flash drives or pamphlets or virtual media such as web pages. There is no direction interaction between the target and the social engineer (Mouton, Leenen, et al., 2014).

5.2.8 Scenario 5

Summary**:** Whilst walking on campus you see a flash drive lying on the ground. It has no identifiable traits on the outside that can be used to identify the owner. You have lost flash drives before and are aware of how much work could be lost that may be saved on the flash drive and feel sorry for whoever may have lost it. What action do you take?

Notable aspects of scenario: You understand how to perform the action; you are capable of performing the action; you do not have the authority to interfere with someone else's property.

Possible responses to scenario:

A. you first scan the flash drive for viruses and if it is found to be virus-free, start examining all folders and opening all files stored on the flash drive to hopefully identify the owner.
B. you decide to install a virtual machine on your computer and use that virtual machine to examine all folders and open all files on the flash drive in an attempt to identify the owner.
C. you give the flash drive to a friend and ask him/her to try identify the owner by examining the files on his/her computer.
D. leave the flash drive where it is, without plugging it into any computer or opening any of the files.

Suggested (i.e., secure) action: Defer or refer request, i.e., choose option C or D.

5.2.9 Response retrieval

To perform the experiment, a Google Forms questionnaire will be used. This questionnaire will present people with the various SE attack scenarios. They are able to select the multiple choice option they feel most accurately depicts how they would react to each scenario. This form of data capture was chosen for its efficiency and ease of use as a link to the questionnaire could be sent out to subjects, with instructions on how to participate in the experiment. Another benefit of this form of data capture is that the results are already in an electronic format, hence reducing the number of errors made during data capture. Furthermore, the results of a Google Forms questionnaire can be exported as .csv file, allowing for easy interpretation of the data using a Python script.

6 Conclusions and Future Work

In conclusion, there is a clear need to develop a tool that can be used in any scenario to determine if the user is being subjected to acts of SE, and to provide guidance as to the correct manner of response to follow in said scenario. The authors have determined that a web implementation of the SEADMv2 model is an effective modus to train individuals in reducing the number of errors made by subjects on various types of scenarios. As such, a methodology and subsequent design of such a web tool is being developed. It is expected to have a significant effect in decreasing the number of errors made on scenarios that employed indirect and bidirectional communication. By executing the envisaged experiment the model efficacy is expected to be validated and also sized. Alongside this web tool, the team has also published work on a mobile implementation of the same model. The results from that research indicates also indicates that a web tool will aid in the prevention of social engineering attacks (Mouton, Teixeira and Meyer, 2017). Future work can also be performed to increase the efficacy of the model in the areas where it was proven to be ineffective by altering the states in the model that deal with aspects unique to scenarios of those types.

7 References

Abraham, S. and Chengalur-Smith, I. (2010) 'An overview of social engineering malware: Trends, tactics, and implications', *Technology in Society*, 32(3), pp. 183–196. doi: http://dx.doi.org/10.1016/j.techsoc.2010.07.001.

Åhlfeldt, R.-M. *et al.* (2005) 'Security Issues in Health Care Process Integration? a Research-in-Progress Report.', in *EMOI-INTEROP*, pp. 1–4.

Bezuidenhout, M., Mouton, F. and Venter, H. S. (2010) 'Social engineering attack detection model: SEADM', in *Information Security for South Africa*. Johannesburg, South Africa, pp. 1–8. doi: 10.1109/ISSA.2010.5588500.

Chantler, A. N. and Broadhurst, R. (2006) *Social engineering and crime prevention in cyberspace*. Queensland University of Technology. Available at: http://eprints.qut.edu.au/7526/1/7526.pdf.

Culpepper, A. M. (2004) *Effectiveness of using red teams to identify maritime security vulnerabilities to terrorist attack*. Naval Postgraduate School.

Gao, W. and Kim, J. (2007) 'Robbing the cradle is like taking candy from a baby', in *Proceedings of the Annual Conference of the Security Policy Institute (GCSPI)*. Amsterdam, Netherlands, pp. 23–37.

Gragg, D. (2002) *A Multi-Level Defense Against Social Engineering*. Available at: http://taupe.free.fr/book/psycho/social engineering/Social Engineering - Sans Institute - Multi Level Defense Against Social Engineering.pdf.

Hadnagy, C. (2010) 'Social Engineering: Past, Present and Future'. (social-engineer.org podcast). Available at: http://www.social-engineer.org/episode-010-social-engineering-past-present-and-future/.

Hamill, J. T., Deckro, R. F. and Jr., J. M. K. (2005) 'Evaluating information assurance strategies', *Decision Support Systems*, 39(3), pp. 463–484. doi: http://dx.doi.org/10.1016/j.dss.2003.11.004.

Ivaturi, K. and Janczewski, L. (2011) 'A Taxonomy for Social Engineering attacks', in Grant, G. (ed.) *International Conference on Information Resources Management*, pp. 1–12.

Ivaturi, K. and Janczewski, L. (2012) 'A Typology Of Social Engineering Attacks ? An Information Science Perspective', in *PACIS 2012 Proceedings*.

Kingsley Ezechi, A. (2011) *Detecting and combating Malware*. University of Debrecen. Available at: http://hdl.handle.net/2437/105305.

Lenkart, J. J. (2011) *The vulnerability of social networking media and the insider threat new eyes for bad guys*. Naval Postgraduate School. Available at: http://calhoun.nps.edu/public/handle/10945/5562.

Van De Merwe, J. and Mouton, F. (2017) 'Mapping the Anatomy of Social Engineering Attacks to the Systems Engineering Life Cycle', in Furnell, S. and Clarke, N. (eds) *Proceedings of the Eleventh International Symposium on Human Aspects of Information Security & Assurance*. Adelaide, Australia, pp. 24–40.

Mitnick, K. D. and Simon, W. L. (2002) *The art of deception: controlling the human element of security*. Edited by W. Publishing. Indianapolis: Wiley Publishing.

Mitnick, K. D. and Simon, W. L. (2005) *The art of intrusion: the real stories behind the exploits of hackers, intruders and deceivers.* Edited by W. Publishing. Indianapolis: Wiley Publishing.

Mouton, F., Malan, M. M., *et al.* (2014) 'Social engineering attack framework', in *Information Security for South Africa*. Johannesburg, South Africa, pp. 1–9. doi: 10.1109/ISSA.2014.6950510.

Mouton, F., Leenen, L., *et al.* (2014) 'Towards an Ontological Model Defining the Social Engineering Domain', in Kimppa, K. et al. (eds) *ICT and Society*. Springer Berlin Heidelberg (IFIP Advances in Information and Communication Technology), pp. 266–279. doi: 10.1007/978-3-662-44208-1_22.

Mouton, F. *et al.* (2017) 'Underlying finite state machine for the social engineering attack detection model', in *2017 Information Security for South Africa (ISSA)*. Johannesburg, South Africa, pp. 98–105. doi: 10.1109/ISSA.2017.8251781.

Mouton, F. *et al.* (2018) 'Finite state machine for the social engineering attack detection model: SEADM', *SAIEE ARJ*, 109(2), pp. 133–147. Available at: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1991-16962018000200004.

Mouton, F., Leenen, L. and Venter, H. S. (2015) 'Social Engineering Attack Detection Model: SEADMv2', in *International Conference on Cyberworlds (CW)*. Visby, Sweden, pp. 216–223. doi: 10.1109/CW.2015.52.

Mouton, F., Leenen, L. and Venter, H. S. (2016) 'Social engineering attack examples, templates and scenarios', *Computers & Security*, 59, pp. 186–209. doi: http://dx.doi.org/10.1016/j.cose.2016.03.004.

Mouton, F., Malan, M. M. and Venter, H. S. (2012) 'Development of cognitive functioning psychological measures for the SEADM', in *Human Aspects of Information Security & Assurance*. Crete, Greece, pp. 40–51.

Mouton, F., Teixeira, M. and Meyer, T. (2017) 'Benchmarking a Mobile Implementation of the Social Engineering Prevention Training Tool', in *Information Security for South Africa*. Johannesburg, South Africa, pp. 106–116. doi: 10.1109/ISSA.2017.8251782.

Nohlberg, M. (2008) *Securing Information Assets: Understanding, Measuring and Protecting against Social Engineering Attacks*. Stockholm University.

Scheeres, J. W. (2008) *Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks*. DTIC Document.

Thornburgh, T. (2004) 'Social engineering: the "Dark Art"', in *Proceedings of the 1st annual conference on Information security curriculum development*. New York, NY, USA: ACM (InfoSecCD '04), pp. 133–135. doi: 10.1145/1059524.1059554.

Workman, M. (2008) 'A test of interventions for security threats from social engineering', *Information Management & Computer Security*. Emerald Group Publishing Limited, 16(5), pp. 463–483.