UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Department of Computer Science

Computer Science Honours
Final Paper
2016

Title: Social Engineering Prevention Training Tool - Testing a Web Implementation of the SEADMv2

Author: Michael Pepper

Project abbreviation: SEPTT

Supervisor: Tommie Meyer

Category	**Min**	**Max**	**Chosen**
Requirement Analysis and Design	0	20	-
Theoretical Analysis	0	25	-
Experiment Design and Execution	0	20	20
System Development and Implementation	0	15	5
Results, Findings and Conclusion	10	20	20
Aim Formulation and Background Work	10	15	15
Quality of Paper Writing and Presentation	10		10
Quality of Deliverables	10		10
Overall General Project Evaluation	0	10	
Total marks			80

Social Engineering Prevention Training Tool - Testing a Web Implementation of the SEADMv2

Michael Pepper
Department of Computer Science
University of Cape Town
mikejpepper@gmail.com

ABSTRACT

The information people possess is often of great value and when stored electronically, is guarded by complicated security mechanisms. These mechanisms are constantly upgraded to counter-act threats that aim at obtaining the information they guard. For this reason, the Social Engineer explores a different avenue of attack and exploits the new 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 SEPTT project addresses this gap by developing a tool that can be used in any scenario to determine if the user is being subjected to acts of Social Engineering, and the correct manner of response to take in said scenario.

In our experiments, the resulting tool indicated a significant reduction in the number of errors made by subjects on potential social engineering attack scenarios, compared to when only intuition was used. The tool also significantly reduced the number of instances where a subject fell victim to genuine social engineering attacks, while not adversely affecting the number of harmless scenarios that were satisfied. Lastly the tool proved most effective at preventing social engineering attacks that were indirect and bidirectionally communicated, while it had no significant effect on unidirectionally communicated attacks.

CCS Concepts

•**Security and privacy** → *Social aspects of security and privacy;*

Keywords

Social Engineering; Attack Prevention; Social Engineering Attack Detection Model; SEADMv2

This thesis was completed as part of a Bachelor of Science Honours degree in Computer Science at the University of Cape Town in 2016. The supplementary material referenced in the paper can be accessed via UCT's Department of Computer Science online publications page at: http://pubs.cs.uct.ac.za. The project abbreviation used for this paper is "SEPTT". All references to appendices made in this paper refer those found on the publications page, as well as the raw data from the experiment.

1. INTRODUCTION

The field of information security is a highly volatile discipline, with the protection of personal information being of vital importance [11]. Hackers are constantly seeking out new ways to exploit different aspects of computer systems [1], with one goal being the retrieval of sensitive personal information. To counter-act this, technological safeguards are developed, ideally mitigating the possibility and impact of such threats. This is a continuous cycle, leading to future attacks being more complicated and having to explore different avenues of attack. Furthermore, organisations, governments and individuals are becoming increasingly aware of the threat of such technology-based attacks and are hence investing in better security technologies [1]. For this reason, some attackers (social engineers) have shifted their focus to exploit the new weakest link in the information security system - the user [11], [12]. This is achieved through the use of psychological ploys which compromise the user's emotional state, hence allowing an exploit to take place [2], [9], [11]. 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 can elicit responses and hence perform information gathering that would not be possible had the user been in a more stable state of mind [12], [2]. 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[10]. This ignorance may be due to their lack of knowledge of the potential gains an attacker can attain from the information they possess. In the USA alone this has led to $3.2 billion in losses due to phishing[1] attacks from August 2006 to August 2007 [13]. Individuals may also have the mind-set that the information in their possession is not of value to anyone, so why should they attempt to protect it [10]? Some individuals also feel they would be able to detect potential social engineering attacks. However this is not the case as in 2004, 1 in 3 people were deemed likely to fall victim to acts of social engineering in their lifetime by the US Department of Justice [15]. The Social Engineer is hence highly skilled at exploiting human vulnerabilities through the use of psychological triggers, in

[1]The act of sending emails to unsuspecting individuals with the aim of persuading them to divulge sensitive personal information, by masquerading as a reputable party.

order to foil human judgement and obtain information [12]. It is for this reason that substantial work has been performed to understand the workings of social engineering attacks, with the aim of detecting and preventing them.

Currently, there is 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" of things to look out for. The Social Engineering Prevention Training Tool (SEPTT) project aims at addressing this gap by implementing the Social Engineering Attack Detection Model Version 2 (SEADMv2) proposed by Mouton et al. [10] as a web application, in order to determine whether it is effective at successfully differentiating between harmless requests and genuine social engineering attacks. In doing this, the user will be guided to the appropriate action to take in a given scenario, hence reducing the probability of them falling victim to a social engineering attack. The hypotheses for this experiment can be summarised as follows:

- A Web-based implementation of the SEADMv2 will significantly reduce the number of scenarios in which subjects fall victim to acts of social engineering.
- A Web-based implementation of the SEADMv2 will significantly increase the number of scenarios in which harmless requests are satisfied by subjects.

The efficacy of the model was assessed through a two stage experiment, whereby subjects were given 10 scenarios that are possible social engineering attacks, with four possible options of how to respond to each scenario. Subjects have to choose the option that most accurately depicts how they would react in each scenario, without the use of the SEADMv2 (**Stage 1**) and with the web implementation of the SEADMv2 (**Stage 2**).

The results of this experiment indicated that the use of the web tool significantly decreased the overall number of errors made on potential social engineering attacks scenarios. This reduction in errors was caused mainly by a reduction in the number of instances where subjects fell victim to genuine social engineering attacks, rather than an increase in the number of harmless requests that were satisfied. There was a significant decrease in the number of instances where a subject fell victim to attack scenarios that were indirect and bidirectionally communicated, whereas unidirectional attacks were seemingly unaffected. These results verify the model's coverage and prediction capabilities, and indicate that the underlying social engineering attack model that the model was built upon is sufficient at modelling and preventing real-world attacks.

The remainder of this paper is structured as follows: Section 2 reviews the current work related to the experiment. Section 3 analyses the design and implementation of the web application and the scenarios that subjects were tested with. Section 4 outlines the methodology used to perform the experiment as well as the manner in which the resulting data was transformed to prepare it for analysis. Section 5 discusses the ethical and professional concerns of the experiment. Section 6 discusses the results of the experiment. Section 7 discusses the limitations of the paper. Section 8 discusses the conclusions that can be drawn from the paper and Section 9 outlines possible future work.

2. RELATED WORK

This section will analyse the current frameworks available to model social engineering (SE) attacks, with emphasis on the framework proposed by Mitnick et al. [8]. 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. The Social Engineering Attack Detection Model Version 2 (SEADMv2) proposed by Mouton et al. [10] will also be discussed as it forms the base of the experiment.

2.1 Attack Frameworks

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 [8] is pivotal in this regard as it is the most widely accepted SE attack framework, since its phases are consistent across all attack types. The cycle breaks an SE attack down into several phases, each of which contains a predetermined goal. These phases will be discussed below, 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 [11]. This information gathering can take many forms and aims at acquiring information and resources necessary to successfully perform an attack. 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 [11]. 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 [1]. 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 they will be more likely to divulge the requested information to the attacker if there is an existing relationship [11]. 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 occupy an authoritative role [11]. In doing this, the attacker hopes to establish some trust connection with the target [4], which will make that target more susceptible to exploitation within 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 model this is achieved through manipulation of the target's emotional state by preying on the seven psychological vulnerabilities noted by Gragg [5]. They are: strong affect, overloading, reciprocation, deceptive relation-

ship, diffusion of responsibility and moral duty, authority, integrity and consistency [14], [7], [16], [3]. By exploiting these psychological vulnerabilities, the target's emotional state is altered and they become more likely to comply with the attacker's requests for information [11].

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 [8]. 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

Social Engineering attacks can be classified according to the manner in which the communication takes place during the exploit, and the interaction between attacker and target [9]. 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 et al. [9], 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 [1].

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 is defined as when two or more parties partake in a conversation and it can be likened to the communication described by Ivaturi & Janczewski [6]. This communication is often performed over interactive mediums such as e-mail and face-to-face conversations as both parties need to be able to contribute. Unidirectional communication is defined as a conversation between attacker and target however the target is not able to converse with the attacker in a back-and-forth manner. Examples of the mediums used include emails and one-way text messages.

2.3 Social Engineering Attack Detection Model Version 2 (SEADMv2)

The SEADMv2 [10] is the second revision of the model initially proposed by Bezuidenhout et al. [2], and provides users with a state diagram that can be used to determine if they are being subjected to acts of SE, and the appropriate action they should take. It achieves this by asking the user questions about their current scenario, the answers to which determine their transition through the model (seen in *Figure 1*). The model eventually reaches a termination state, indicating to the user whether they should *Perform the Request* or *Defer or Refer Request*. *Perform the Request* suggests to the user that they should comply with the requester's demands and perform the relevant action as it is likely not an SE attack. *Defer or Refer Request* suggests that they may be involved in an SE attack and should refer the request to someone more well-suited to deal with said request, or defer the request completely - whichever is more applicable.

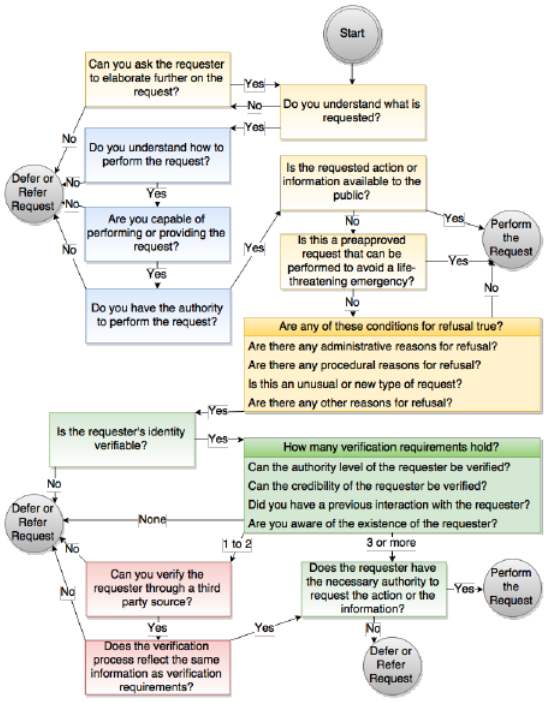

Figure 1: Mouton et al. [10] Social Engineering Attack Detection Model Version 2

This version of the SEADM improves upon the first iteration proposed by Bezuidenhout et al. [2] by expanding upon the states proposed, hence increasing the model's coverage and making it more user friendly. The state in the previous model that required the user to evaluate their emotional state has also been omitted, and is dealt with by a separate psychological measure described by Mouton et al. [12].

3. DESIGN AND IMPLEMENTATION

To perform this experiment, a Web application that allows users to traverse the SEADMv2 in a natural and efficient manner was created. Subjects were then exposed to potential SE attacks both with and without the use of this tool. This section will discuss the design considerations and techniques that were employed to develop this application, as well as the scenarios that were created to assess the tools efficacy. The questionnaire through which the experiment was conducted will also be discussed.

3.1 Web Application

The Web application that was developed (seen in *Figure 2*) consists of a question box that displays the relevant question according to the user's current state in the SEADMv2. This question aims at assessing the user's knowledge of the current situation in order to transition to the next state in the model and eventually determine the correct action to take. Below that, there are two buttons that allow the users to answer these questions, labelled "Yes" and "No" respectively. There is a progress bar on the left side of the interface indicating to the user their current position in the

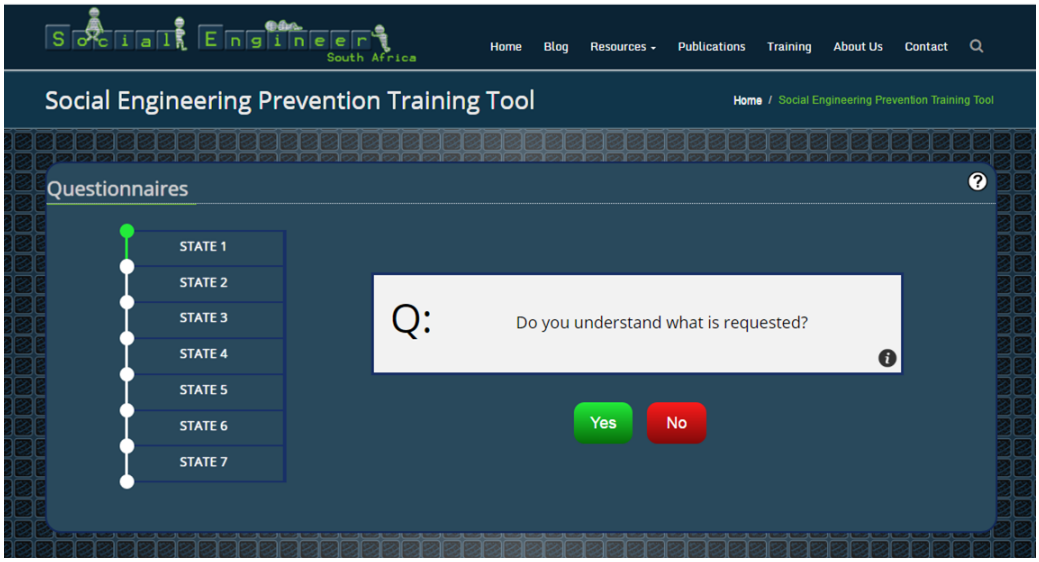

Figure 2: Social Engineering Prevention Training Tool Web Application

model, and informational buttons that can be used to aid the user should they not understand 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 is hosted on www.social-engineering.co.za and makes use of a MySQL database to store the SEADMv2 model.

3.2 Social Engineering Scenarios

Once the Web application was developed, 10 believable real-life situations were drafted into scenario format. These scenarios focused on the *Develop Rapport and Trust* and *Exploit Trust* phases of Mitnick's attack cycle [8], and employed common techniques noted within these phases that aid in the execution of a successful SE attack. Each scenario had four possible answers, each option depicting a different type of response to that scenario. Two of these options depicted responses that complied with the requests in the scenario and performed the relative action, and two of the options did not comply or referred the scenario to someone else.

Since the SEADMv2 indicates to users whether to perform the request in a given scenario or not, scenarios that are "harmless" and not attempted SE attacks were included within the 10 scenarios being tested. This ensures that the model's ability to differentiate between actual SE attacks and harmless requests can be assessed, as if only SE attack scenarios were used, the model could simply indicate to always *Defer or Refer Request* and hence prevent all possible attacks. This would not be useful as genuine requests would never be satisfied, thus negating the model's real-world applicability. The resulting 10 scenarios consisted of 8 genuine attack scenarios and 2 harmless scenarios. After completing the experiment it became clear that a more even split of harmless and attack scenarios would have been ideal as the low number of harmless scenarios affected the credibility of those results. This lack of foresight is discussed in the limitations section and arose during the planning stages of the experiment where the only consideration was that there were harmless scenarios, not necessarily how many. 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 attack classifications described by Mouton et al. [9] were used as templates. Of the 10 scenarios that were created, there were 5 that depicted unidirectional communication, 4 depicted bidirectional communication and 1 depicted indirect communication. All 10 of these scenarios can be found in *Appendix A*. For the sake of conciseness, 5 scenarios that are representative of the 10 created will be discussed below.

3.2.1 *Unidirectional Communication*

Scenario 1.

Summary: Whilst at work you receive an email from a new email address saying that a new person from your company's external accounting firm has started working on the time reports for this quarter and hence needs you to send your preliminary time report through as soon as possible. The email address that the message came 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; Their identity, authority and credibility is verifiable; You have had no previous interaction with the requester but can verify their intentions.

Possible Responses to scenario:

A) Since you don't 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) 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) Contact your superior to find out whether or not they approve of you sending your preliminary time report to the person requesting it or not.

D) Refuse to send her your preliminary time report.

Suggested Action: *Perform the Request* - Option A or Option B

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 7 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) Provide all the requested information.

B) Ask the guest lecturer a few complementary questions and based on his answers decide whether to provide the information.

C) Ask the guest lecturer to rather contact your lecturer directly to obtain this information.

D) Do not provide the information and also don't tell the guest lecturer where to get it as I deem it to be sensitive information.

Suggested Action: *Defer or Refer Request* - Option C or Option D

3.2.2 Bidirectional Communication

Scenario 3.

Summary: You receive a message on Facebook from a random person claiming to be 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 on to 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 & 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; Their identity is not verifiable.

Possible Responses to scenario:

A) 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) 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) 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 on Facebook Messenger any of your details.

D) 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's best not to enter.

Suggested Action: *Defer or Refer Request* - Option C or Option D

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, however you can see that the student is stressed as they have 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) 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 them in or not.

C) Refuse to help the student at all and tell them they shouldn't have waited till the last minute to submit their assignment and they should always have their student card on them while on campus.

D) Give the student directions to the access control offices where the student can prove their identity and hopefully get access to a computer lab within 15 minutes to submit their assignment.

Suggested Action: *Defer or Refer Request* - Option C or Option D

3.2.3 Indirect Communication

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) Give the flash drive to a friend and ask them to try identify the owner by examining the files on their computer.

D) Leave the flash drive where it is, without plugging it into any computer or opening any of the files.

Suggested Action: *Defer or Refer Request* - Option C or Option D

3.3 Response Retrieval

To perform the experiment a Google Questionnaire was used (available at https://goo.gl/forms/KCIjN3Dx4apYXuZ22). This questionnaire presents people with the various SE attack scenarios and 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 partake 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 the Google Questionnaire can be exported as .csv file, allowing for easy interpretation of the data easy with Python code.

4. METHODOLOGY

To test the hypotheses of the paper, a two stage experiment was run with 45 subjects. These subjects volunteered to take part in the study and were composed of 39 university students and 6 high school teachers. Subjects were sent a link to a Google Questionnaire and instructed that they would be partaking in a two stage experiment and how to go about completing the experiment. The order in which questions were asked in both stages was randomised so as to avoid any ordering effects on subjects' answers. These 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.

4.1 Experiment

Stage 1 : The first stage consisted of the 10 potential SE attack scenarios described in *Section 3.2*, each with four possible answers. Subjects were told to select one of the four possible answers, according to which answer best represented how they would react to that scenario in real-life. This answer is based purely on gut feeling and results in a record of how subjects would respond to each scenario without any assistance. This forms the "Without Model" before-treatment data and is the control of the experiment.

Stage 2 : Upon completion of *Stage 1*, subjects were informed that they must now make use of the web implementation of SEADMv2 model to guide their answers to the previous 10 scenarios. To achieve this, the same 10 scenarios were 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 asks. Once a subject reaches a terminating state in the model, it will indicate that they should either "Perform the Request" or "Defer or Refer the Request". The subject must use this information to select the multiple choice option that complies with that guidance. For example if the model indicates that for a specific scenario the user should "Defer or Refer the Request", the subject must choose a multiple choice option that does not comply with the requests in the scenario, or defers the situation to someone more well-equipped to deal with it. The result of this stage of the experiment is 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 were limited to one per person to prevent the same person answering it multiple times and skewing the data. Since the SEADMv2 model aims at changing the way you assess any given scenario by giving you questions to consider, it is obvious that this would pose an issue. To incentivise participation in the experiment, it was advertised that 5 prizes worth R200 would be randomly drawn amongst the participants. The winning contestants were subsequently notified by emailing the email address they used to answer the Google Questionnaire. This random draw was performed 2 weeks after all results had been recorded, to prevent any biases from entering the subject's answers.

4.2 Data Transformation

Once the experiment was completed, the data obtained needed to be transformed into a more analysis-friendly format. Firstly, the data contained each subject's answers to the 10 scenarios both with and without the use of the web application (20 data points per subject). Naturally these answers were very long so they were transformed to be either A, B, C or D, according to which option they correlated to in the relative scenarios. Once transformed to this easier to

use format, these answers were compared against a model solution. This resulted in a binary indication of whether the subject chose a correct option for a given scenario, or if they chose an incorrect option. It is worth noting that this binary indicator does not necessarily indicate that the subject fell victim to an SE attack in that scenario as not all scenarios represent SE attacks. This indicator merely documents whether the correct reaction to a scenario was chosen and when summed across all scenarios indicates the total number of incorrect reactions by a subject. The result of this transformation was a binary record of whether a subject chose a correct answer to each scenario without the model, and with the model.

5. PROFESSIONAL AND ETHICAL ISSUES

Ethical clearance was obtained from the Science Faculty Research Ethics Committee and the Department of Student Affairs. Before partaking in the experiment, participants signed a consent form indicating that their results would be anonymous and that they would only be used for the purpose of this experiment.

No alterations were performed to the experimental data once it had been obtained. From a professional point of view this ensures that the results can safely be shared with the community, without misrepresenting any findings.

6. RESULTS

The overall results of the experiment have been summarised in *Table 1*, showing the total number of incorrect answers with and without the model; the total number of answers that were changed from wrong to right for a given scenario through the use of the model; and the total number of answers that remained wrong / right even with the use of the model. For the remainder of this section an incorrect option choice for a given scenario will be referred to as an "error". The statistical analysis to follow was performed using RStudio.

Total Errors Without Model	190
Total Errors With Model	149
Total Wrong to Right	97
Total Right to Wrong	56
Total Stayed Wrong	93
Total Stayed Right	204

Table 1: Totals of Different Scenario Phenomenon

6.1 Overall Model Efficacy

Upon inspection of these results, one can see that overall there were more errors made without the use of the model (190) than with the use of the model (149). Since errors are usually Poisson distributed, it follows that the underlying datasets for these totals should follow that distribution too. If this is the case, one can compare the means of those two datasets and determine if there is in fact a significant difference between them. To test if these datasets follow a Poisson distribution, the number of errors made by each subject were calculated both with and without the use of the model and a Chi-Squared goodness of fit test was run on that data. These tests indicated that the number of errors made by each subject without the model were not Poisson distributed ($\mathbf{p = 0.04}$), while the number of errors made with the use of the model were not significantly different from a Poisson distribution ($\mathbf{p = 0.39}$). As both datasets did not follow a Poisson distribution we used an alternative test to that initially planned to compare their means.

Since there is only one within-subjects factor with two levels (Without Model, With Model) in this experiment, one can run a Wilcoxon signed-rank test to determine if their means are significantly different. This non-parametric test is not as powerful as the parametric equivalent, but since the two datasets follow different distributions, this is a necessary compromise.

This test indicated that there was a significant difference between the mean number of errors made per subject with and without the use of the model ($\mathbf{p < 0.01}$). This difference suggests that the web implementation of the SEADMv2 model had a significant impact in reducing the overall number of scenarios in which a subject reacted incorrectly. This is reinforced by the data represented in *Figure 3* where it is clearly indicated that the means of these two datasets are noticeably different, with the "Without Model" plot being noticeably higher.

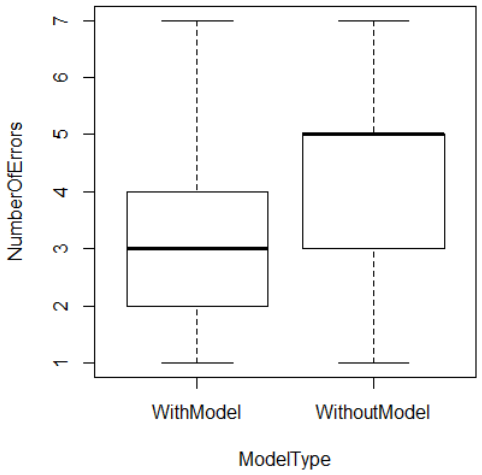

Figure 3: Box Plot Indicating Number of Errors Made By Subjects With and Without The Model

For the remainder of the analysis, the Bonferroni correction will be used in order to counteract the problem of multiple comparisons within a dataset. This correction reduces the chance of incorrectly rejecting the null hypothesis of a given test (type 1 statistical error) by testing each hypothesis at a significance level of α/κ where α is the desired overall significance level and κ is the number of hypotheses being tested. This accounts for the increasing probability of type 1 errors inherent with multiple comparisons.

6.2 Model Success on Different Threat Scenarios

Once it was established that the model was effective at an overall level, testing began to determine if the individual hypotheses were correct. The number of errors per subject were first categorised according to the threat-level of the scenario that they arose from, being either "harmless" or "attack". This resulted in a record of the number of errors made by each subject with and without the model, for both

genuine attacks and harmless scenarios. *Table 2* outlines the general format of this data, the analysis of which can be found in *Section 6.2.1*. For the analysis in this section two hypotheses are being tested, with a Bonferroni corrected p-value as follows:

p-value: $\alpha/\kappa = 0.05/2 = 0.025$.

Subject No.	Model Type	Scenario Threat-Level
1	Without	Harmless
1	Without	Attack
1	With	Harmless
1	With	Attack

Table 2: Breakdown of Errors Made by Subjects According to Use of Model and Scenario Threat-Level

6.2.1 Attack Scenario Prevention

To test the first hypothesis, a Chi-Squared goodness of fit test was performed on the data detailing the number of errors made per subject on genuine attack scenarios, with and without the model. This test determined that the distribution of the errors made without the use of the model was not significantly different from a Poisson distribution at the 5% level, once the Bonferroni adjustment had been accounted for ($\mathbf{p = 0.046 > 0.025}$). This was also the case for the errors made with the use of the model, however this test was much more definitive ($\mathbf{p = 0.24}$). Since there was such a close margin of potentially concluding the first distribution was not Poisson distributed, it is more applicable to run a Wilcoxon signed-rank test to compare their means. This will yield more credible results than the parametric equivalent in this case as the distributions are so variable. This test indicated that there was a significant difference in the mean number of errors made by subjects with and without the use of the model ($\mathbf{p = 0.002}$). This is reinforced by the data illustrated in *Figure 4* where it is clearly indicated that the means of these two datasets is noticeably different, again with the "Without Model" plot being significantly higher.

We can therefore conclude that the use of the model significantly decreased the number of errors made by subjects in genuine attack scenarios and hence prove the first hypothesis of the paper. It follows therefore that the use of the model significantly reduced the number of instances that a subject fell victim to social engineering attacks, as on average they chose answers that would prevent the attack from being successful more often with the model than without it. This reduction in the number of instances that subjects fell victim to attacks pays testament to the model's ability to model any scenario and accurately determine if it is an SE attack, hence preventing the attack.

6.2.2 Harmless Scenario Compliance

The second hypothesis was tested in much the same manner as above, firstly analysing whether the number of errors made per subject on harmless scenarios, with and without the use of the model, were Poisson distributed. This turned out not to be the case for both the "With Model" and "Without Model" data as the Chi-Squared goodness of fit tests returned p-values that indicate to reject the null hypothesis ($\mathbf{p < 0.01}$). A Wilcoxon signed-rank test was hence run on the data, indicating that there was not a significant difference

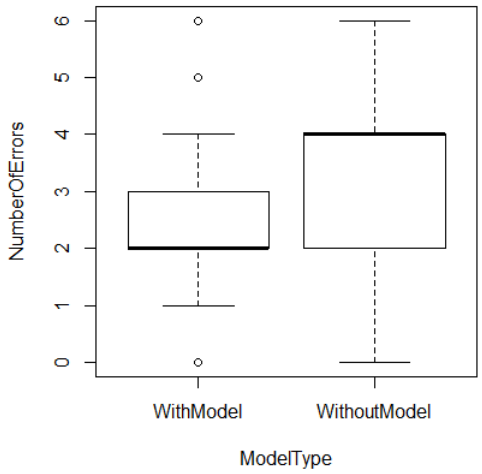

Figure 4: Box Plot Indicating Number of Errors Made by Subjects on Genuine Attack Scenarios, With and Without The Model

between the means of these two datasets ($\mathbf{p = 0.87}$). This is depicted in *Figure 5* where it is clear that the number of subjects that fall into each error count category does not seem to change significantly with the use of the model.

It is worth noting that these results are somewhat limited as there were only 2 "harmless" scenarios and hence there was not a large spread of number of errors that could be made on such attacks (ranging from 0 - 2). This means that the averages of the two datasets is not likely to be vastly different. However the similarity in the histograms of the error counts before and after the model gives us confidence that the correct conclusion has been reached.

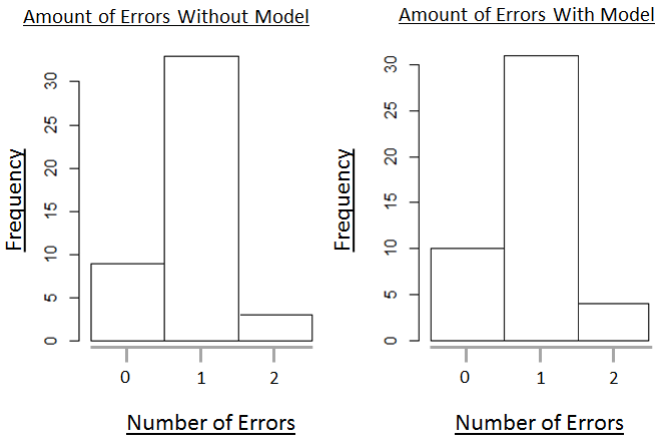

Figure 5: Histograms Showing Frequency of Subjects According To Number of Errors Made

We hence conclude that the use of the model did not significantly decrease the number of errors made by subjects on harmless scenarios. This suggests that the model does not significantly influence the number of harmless scenarios in which a subject complies with the requester's demands or performs the required action. We therefore reject the second hypothesis of the paper.

In review, *Section 6.1* established that the model was effective at an overall level at reducing the number of errors made by subjects. We have now shown that the number of errors made on genuine attacks was also significantly decreased through the use of the model, however the number made on harmless scenarios was not. The data in *Figure 6* illustrates this point as it is clearly indicated that the significant decrease in the overall number of errors made by subjects can be almost entirely attributed to the reduction of errors in genuine attack scenarios. This reinforces the efficacy of the model as there are significantly fewer cases where subjects fall victim to SE attacks with the use of the model, while the number of harmless requests satisfied is mostly unaffected. In an ideal situation, the number of harmless requests satisfied would increase when using the model, however this is not the case. We view this as an acceptable shortfall of the model, as subjects are less susceptible to SE attacks with the model and equally as compliant to harmless requests.

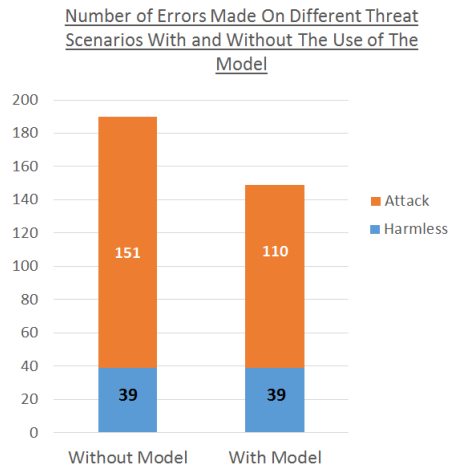

Figure 6: Histograms Showing Count of Subjects According To Number of Errors Made

6.3 Direct and Indirect Attacks Statistics

The errors per subject were then divided according to the communication used in the scenario that they arose from, being either "Unidirectional", "Bidirectional" or "Indirect". *Table 3* outlines the general format of this data, the analysis of which follows, using a Bonferroni corrected p-value $= \alpha/\kappa = 0.05/3 = 0.017$. A κ value of 3 was used in this correction since for each communication type, the hypothesis being tested was that there was no significance different between the mean number of errors with and without the use of the model and hence 3 hypotheses were being tested.

6.3.1 Unidirectional

Firstly, the number of errors made per subject on unidirectional scenarios with and without the use of the model were tested. The same tests as in *Section 6.2* were performed indicating that the number of errors without the model were not Poisson distributed ($\mathbf{p = 0.008}$), while with the model we failed to reject the null hypothesis (just) and concluded they were Poisson distributed ($\mathbf{p = 0.04}$). For reasons stated earlier, a Wilcoxon signed-rank test was then performed, indicating that the means of the two datasets were not significantly different. Given this information we can conclude that the use of the model had no significant effect on reducing the number of errors made on scenarios that employed unidirectional communication.

Subject No.	Model Type	Attack Communication
1	Without	Unidirectional
1	Without	Bidirectional
1	Without	Indirect
1	With	Unidirectional
1	With	Bidirectional
1	With	Indirect

Table 3: Breakdown of Errors Made By Subjects According To Use of Model and Scenario Threat-Level

6.3.2 Bidirectional

The same tests as above were run on the Bidirectional data, however the Chi-Squared goodness of fit test indicated that the distributions of both the "Without Model" data ($\mathbf{p = 0.67}$) and the "With Model" data ($\mathbf{p = 0.52}$) were not significantly different from a Poisson distribution. We hence used an exact Poisson test to determine if the mean error rates (λ) of the two distributions were different, where "Without Model" $\lambda = 1.489$; "With Model" $\lambda = 0.7778$; sample size $= 45$. This test indicated that the two λ values were significantly different ($\mathbf{p = 0.0025}$), and we hence conclude that the model was effective at reducing the number of errors made by subjects on scenarios that employed bidirectional communication.

6.3.3 Indirect

Lastly the data pertaining to indirect scenarios was assessed, indicating that the distribution of the "Without Model" dataset was significantly different from a Poisson distribution ($\mathbf{p = 0.001}$), while the "With Model" dataset was not ($\mathbf{p = 0.034}$). The Wilcoxon signed-rank test run on this data indicated that their means were significantly different ($\mathbf{p = 0.012}$). It is worth noting that there was only one scenario that fell into this category and hence the applicability of the results to all indirect communication based attacks is limited. However we can conclude that in this case, the use of the model significantly decreased the number of errors made on scenarios that employed indirect communication. The model hence reduced the probability of subjects falling victim to such attacks.

6.4 Notable Results

This final stage of the analysis refers to *Table 1* which indicates that there were 97 instances where a subject's answer changed from incorrect to correct through the use of the model, while there were 56 instances where a correct answer was changed to incorrect. There were also 93 instances where a subject's answer remained incorrect even with the use of the model, while 203 remained correct.

From this information we can determine that subjects who answered a scenario correctly without the model, answered the same scenario incorrectly 22% of the time with the model. This phenomenon is naturally undesirable, however it is outweighed by the 51% of incorrect answers that were corrected through the use of the model. Furthermore

it is apparent that the number of answers that remained correct with the use of the model (204) is much higher than the number that remained incorrect(93). When these statistics are combined, we conclude that the model is effective at reducing the number of errors made by subjects, whilst ensuring that correct answers remain correct more often that not.

A possible explanation for the number of instances where a subjects answer was changed from correct to incorrect through the use of the model is the confusing wording of the states in the model. Users often voiced feelings of confusion when using the model as they found it difficult to understand how to relate questions in the model to a given scenario. This may have led them to answer that question incorrectly and hence transition through the model incorrectly. This could have resulted in the reaching of an incorrect termination state in the model which would guide them to select an incorrect response to the scenario. In such cases, a correct answer would be negatively altered through the use of the model as a result of its abstract wording, rather than its prediction capabilities. This is explained further in the limitations section of the paper.

7. LIMITATIONS

During the initial planning stages of this experiment there was a lack of foresight into the statistical tests that would be run on the resulting data. For this reason, the scenarios that were created were not equally distributed amongst the different communication classifications, and threat levels. This resulted in the statistical tests relating to scenarios that used indirect communication and scenarios that were harmless having less credibility than is ideal.

The overall number of the participants in the study was also greatly affected by the "Fees Must Fall" protests that were occurring during the time of testing, and hence the sample size was substantially smaller than initially planned. This naturally affects the credibility of the overall results. However this was an uncontrollable compromise.

A common grievance amongst subjects during the experiment was that the wording of the questions in the different states of the SEADMv2 framework was very difficult to understand and relate to the scenario being tested. Subjects often voiced feelings of confusion and frustration as the wording of certain questions was too abstract and generalised to apply to a given scenario and hence they felt unsure how to proceed. This may have lead to incorrect transitions through the model and subsequently the incorrect result may have been obtained.

Lastly a general lack of statistical know-how possessed by the SEPTT team may have lead to minor statistical errors in testing. Attempts to mitigate this were made by consulting with Dr. Brian DeRenzi on multiple occasions. Unfortunately the Statistical Sciences Department at the University of Cape Town were unable to provide assistance within the required time frame and hence online resources were used *ad nauseam*.

8. CONCLUSIONS

In conclusion, we have determined that a web implementation of the SEADMv2 model is effective at reducing the number of errors made by subjects on various types of scenarios. Subsequently the model is effective at significantly reducing the number of errors made by subjects on genuine attack scenarios, and hence when used reduces the probability of a subject falling victim to SE attacks. The model was proven to have no significant effect on increasing the number of harmless scenarios that were performed / complied with. It was hence concluded that the changes observed in the overall number of errors made by subjects with and without the model can be almost entirely attributed to the prevention of actual SE attack scenarios. The first hypothesis of the paper was hence proved, while the second hypothesis was rejected.

The tool was proven to have a significant effect in decreasing the number of errors made on scenarios that employed indirect and bidirectional communication, and hence when used subjects are significantly less likely to fall victim to those types of SE attacks than when the model is not used. The model was proven to have no significant effect in preventing scenarios that were unidirectionally communicated.

Lastly the model was shown to correct subjects' answers more commonly than it made them incorrect. There was also a much higher rate of answers remaining correct with the use of the model than the amount that remained incorrect. We hence concluded that the model is effective at reducing the number of errors made by subjects, whilst ensuring that correct answers remain correct more often that not.

Overall the benefits of the model proved statistically significant, while its only downfall was a lack of efficacy at increasing how many harmless scenarios were satisfied.

9. FUTURE WORK

This paper has made a considerable contribution to the field of social engineering, extending the work on the SEADMv2 framework by testing its efficacy. In the future, work can be done to attempt to re-word the model so as to make it more user friendly. This may increase its efficacy by making it easier to understand and relate to a given scenario, without having to change the underlying framework.

Work can also be performed to increase the efficacy of the model in the areas where it was proven to be ineffective (unidirectional scenarios, harmless scenarios etc.) by altering the states in the model that deal with aspects unique to scenarios of those types. For example in unidirectional scenarios, users often did not understand how to verify the identity of a party they could not converse with. To remedy this, the identity verification state could be divided into sub-states that ask more specific verification questions e.g. Does a Google search verify the requester's identity; Does a review website have a record of the company etc.. While this does not alter the states of the model explicitly, it does make them more user friendly and hence may increase the efficacy of the model.

10. ACKNOWLEDGEMENTS

Special thanks goes to Francois Mouton and Tommie Meyer for their constant guidance and input throughout the project. The result would not have been the same had you not been involved.

Thank you also to Michelle Kuttel for her understanding of the unique pressures faced by the honours group this year. It was greatly appreciated.

11. REFERENCES

[1] Abraham, S., and Chengalur-Smith, I. An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society 32*, 3 (2010), 183–196.

[2] Bezuidenhout, M., Mouton, F., and Venter, H. S. Social engineering attack detection model: Seadm. In *Information Security for South Africa (ISSA), 2010* (2010), IEEE, pp. 1–8.

[3] Chantler, A. N., and Broadhurst, R. Social engineering and crime prevention in cyberspace.

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

[5] Gragg, D. A multi-layer defense against social engineering. *SANS Insitute Reading Room* (2002).

[6] Ivaturi, K., and Janczewski, L. A taxonomy for social engineering attacks. In *International Conference on Information Resources Management, Centre for Information Technology, Organizations, and People (June 2011)* (2011).

[7] Mitnick, K. D., and Simon, W. L. *The Art of Intrusion: The real stories behind the exploits of hackers, intruders and deceivers*. John Wiley & Sons, 2009.

[8] Mitnick, K. D., and Simon, W. L. *The art of deception: Controlling the human element of security*. John Wiley & Sons, 2011.

[9] Mouton, F., Leenen, L., Malan, M. M., and Venter, H. Towards an ontological model defining the social engineering domain. In *ICT and Society*. Springer, 2014, pp. 266–279.

[10] Mouton, F., Leenen, L., and Venter, H. Social engineering attack detection model: Seadmv2. In *2015 International Conference on Cyberworlds (CW)* (2015), IEEE, pp. 216–223.

[11] Mouton, F., Malan, M. M., Leenen, L., and Venter, H. S. Social engineering attack framework. In *Information Security for South Africa (ISSA), 2014* (2014), IEEE, pp. 1–9.

[12] Mouton, F., Malan, M. M., and Venter, H. S. Development of cognitive functioning psychological measures for the seadm. In *HAISA* (2012), pp. 40–51.

[13] Sandouka, H., Cullen, A., and Mann, I. Social engineering detection using neural networks. In *CyberWorlds, 2009. CW'09. International Conference on* (2009), IEEE, pp. 273–278.

[14] Scheeres, J. W. Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks. Tech. rep., DTIC Document, 2008.

[15] Workman, M. Gaining access with social engineering: An empirical study of the threat. *Information Systems Security 16*, 6 (2007), 315–331.

[16] Workman, M. A test of interventions for security threats from social engineering. *Information Management & Computer Security 16*, 5 (2008), 463–483.

Social Engineering Prevention Training Tool - Appendix A

Michael Pepper
Department of Computer Science
University of Cape Town
mikejpeppe@gmail.com

1. SOCIAL ENGINEERING SCENARIOS

All 10 of the scenarios used in the experiment of the SEPTT project can be found below.

1.1 Unidirectional Communication

Scenario 1 .

Summary: As a student looking for a job, a job recruiter connects with you on LinkedIn and asks to get in contact so that he can find you a job and receive a commission fee. The recruiter asks a few basic questions about your plans for work and about your interests. Contact details for the recruiter and the recruitment company are attached in the message, as well as a link to the company's website. What action do you take?

Notable aspects of scenario: Requester is a complete stranger; His identity is verifiable; Information requested is not sensitive.

Possible Responses to scenario:

A) You have been job hunting for ages and see this as a blessing in disguise. You reply to the email saying that you would appreciate his help and attach a copy of your CV, a copy of your ID and a link to your GitHub account (since you know he will most likely request all this information anyway).

B) Give the recruiter a call on one of the two numbers provided and arrange to meet up with him to discuss how he can help you.

C) Click on the links provided, and have a look at the company's online profile as well as reviews left by others. If all seems good, get in contact with the recruiter.

D) Ignore the email and remove the recruiter from your LinkedIn.

Suggested Action: *Perform the Request* - Option B or Option C

Scenario 2.

Summary: Whilst at work you receive an email from a new email address saying that a new person from your company's external accounting firm has started working on the time reports for this quarter and hence needs you to send your preliminary time report through as soon as possible. The email address that the message came from does have 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: Requester is a complete stranger; Their identity and credibility is verifiable; 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.

Possible Responses to scenario:

A) Since you don't 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) 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) Contact your superior to find out whether or not they approve of you sending your preliminary time report to the person requesting it or not.

D) Refuse to send her your preliminary time report.

Suggested Action: *Perform the Request* - Option A or Option B

Scenario 3.

Summary: You are working in the computer labs with a friend to finish an assignment that is due in 30 minutes. You receive an email from your lecturer that gave you the assignment, stating that there is a different submission link for the assignment than previously stated. You are fairly certain that this email is from your lecturer, however it is uncommon for lecturers to email students directly and not use the announcement system on the course's website. Which action do you take?

Notable aspects of scenario: 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 reply to the lecturer's email asking a few complementary questions, before accepting the new link to be correct and submitting to it.

B) You ask your friend whether they also received the email and based on your friend's answer decide where to submit your assignment.

C) Submit the assignment to the new link, without asking any questions.

D) Deny the new submission link and use the old (original) submission link instead.

Suggested Action: *Defer or Refer Request* - Option B or Option D

Scenario 4.

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 7 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: 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) Provide all the requested information.

B) Ask the guest lecturer a few complementary questions and based on his answers decide whether to provide the information.

C) Ask the guest lecturer to rather contact your lecturer directly to obtain this information.

D) Do not provide the information and also don't tell the guest lecturer where to get it as I deem it to be sensitive information.

Suggested Action: *Defer or Refer Request* - Option C or Option D

Scenario 5.

Summary: A day before an assignment is due you receive an email from a fellow student. You have never met this student before, but in the email the student says he is in the same class as you and that he got your email address from the participants section on the course's Vula tab. Attached to the email is a PDF of that student's assignment. The student says that he is not sure if his answers are correct and he would really appreciate your help. He asks you to please read through his assignment, compare his answers to yours and then let him know if there are any differences. You open the attached PDF and can clearly see that the student has done the assignment. Which action do you take?

Notable aspects of scenario: 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 are very busy finishing off your assignment and don't have time to look at his answers. You therefore decide to just send him your assignment so far and say he is welcome to compare answers, but you don't have time.

B) You see this as a life saver and decide to use this to your advantage by comparing his answers to yours and using his answers to help you finish your assignment. You then also reply to his email telling him where your answers were different to his.

C) You see this as a life saver and decide to use this to your advantage by comparing his answers to yours and using his answers to help you finish your assignment. You don't respond to his email, however, since you do not know him and wouldn't want his answers to be identical to yours out of fear of being caught for plagiarism.

D) You close the PDF immediately after opening it and delete it. You don't respond to the email and decide to complete your assignment on your own without comparing your answers to the answers sent to you by the student.

Suggested Action: *Defer or Refer Request* - Option C or Option D

1.2 Bidirectional Communication

Scenario 6.

Summary: Whilst on vacation, a lady claiming to be a newly employed technician of one of your customers phones you and notifies you that they will be updating the backup system and hence needs to know where the files you worked on can be found. She asks further for the computer the files can be located on as well as your username in order to ensure the files will be backed up in the new system. You comply with these requests and give her the information required.

The lad then phones back in an hour stating that something went wrong and only your files are causing trouble with the backup. She asks if you would come into the office to sort it out - something you refuse. Since you will not come in to the office she asks for your login information so that she can check that the files have not been destroyed and that everything is okay. What action do you take?

Notable aspects of scenario: Requester is a complete stranger; Her identity is not verifiable; 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

Possible Responses to scenario:

A) Instantly provide her with your login details and wish her a merry Christmas.

B) Ask her a few complementary questions and based on her answers either provide her with your login details or don't provide her with any information.

C) Ask her to contact one of your colleagues since you are on holiday.

D) Refuse to help her entirely and hang up the phone.

Suggested Action: *Defer or Refer Request* - Option C or Option D

Scenario 7.

Summary: You receive a message on Facebook from a random person claiming to be 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 on to 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 & password) since an entry requires a link to your Facebook account. What action do you take?

Notable aspects of scenario: Requester is a complete stranger; Their identity is not verifiable; 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.

Possible Responses to scenario:

A) 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) 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) 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 on Facebook Messenger any of your details.

D) 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's best not to enter.

Suggested Action: *Defer or Refer Request* - Option C or Option D

Scenario 8.

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, however you can see that the student is stressed as they have an assignment to submit within the next 15 minutes. What action do you take?

Notable aspects of scenario: Requester is a complete stranger; Their identity is not verifiable; You do not have the authority to perform this request.

Possible Responses to scenario:

A) 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 them in or not.

C) Refuse to help the student at all and tell them they shouldn't have waited till the last minute to submit their assignment and they should always have their student card on them while on campus.

D) Give the student directions to the access control offices where the student can prove their identity and hopefully get access to a computer lab within 15 minutes to submit their assignment.

Suggested Action: *Defer or Refer Request* - Option C or Option D

Scenario 9.

Summary: You have been working on a major project for a large customer for a long time. Within the project you use special software that was designed specifically for the project, however you do not know too much about its inner workings. You receive a phone call from a man claiming to be from technical support. He explains that you seem to be running an old version of the software, something that has given them severe problems as it hinders the encryption from working correctly. The man directs you to a link where a more recent version can be found and asks you to install it as soon as possible and to log in again to stop the problems. What action do you take?

Notable aspects of scenario: Requester is a complete stranger; Their identity is not verifiable; You have the authority to perform the request; The information involved is not publicly available; This is not a pre-approved request; There are procedural reasons for refusal.

Possible Responses to scenario:

A) I instantly comply and follow the new download link.

B) I ask a number of complementary questions before I comply by clicking the download link.

C) I ask complementary questions and ask if I can come back to him at a later stage.

D) I do not comply with his request at all.

Suggested Action: *Defer or Refer Request* - Option C or Option D

1.3 Indirect Communication

Scenario 10.

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: The owners identity is not verifiable; 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) Give the flash drive to a friend and ask them to try identify the owner by examining the files on their computer.

D) Leave the flash drive where it is, without plugging it into any computer or opening any of the files.

Suggested Action: *Defer or Refer Request* - Option D or Option C